













FILE COPY

DO NOT REMOVE

NBSIR 77-1276

# OMNITAB II User's Reference Manual 1977 Supplement

David Hogben  
Sally T. Peavy

RECEIVED  
DATE 8/18/77  
OTP

Institute for Basic Standards  
National Bureau of Standards  
Washington, D.C. 20234

July 1977

Final

Prepared for  
National Highway Traffic Safety Administration  
Department of Transportation  
Washington, D.C. 20590



NBSIR 77-1276

# **OMNITAB II USER'S REFERENCE MANUAL 1977 SUPPLEMENT**

---

David Hogben  
Sally T. Peavy

Institute for Basic Standards  
National Bureau of Standards  
Washington, D.C. 20234

July 1977

Final

Prepared for  
National Highway Traffic Safety Administration  
Department of Transportation  
Washington, D.C. 20590



---

**U.S. DEPARTMENT OF COMMERCE, Juanita M. Kreps, *Secretary***

**Dr. Sidney Harman, *Under Secretary***

**Jordan J. Baruch, *Assistant Secretary for Science and Technology***

**NATIONAL BUREAU OF STANDARDS, Ernest Ambler, *Acting Director***



OMNITAB II User's Reference Manual  
1977 SUPPLEMENT

by

David Hogben and Sally T. Peavy

Preface

In the early 1960's Joseph Hilsenrath designed a totally new general purpose numerical scientific and statistical language known as OMNITAB. The heart of the OMNITAB design was the universal worksheet concept and the use of very simple English like instructions to perform scientific calculations on data in columns of the worksheet. With the help of numerous colleagues, the language was developed to the point where it became a very effective and easy to use tool for the analysis of data.

The first full documentation appeared in NBS Handbook 101. NBS Handbook 101 was published in 1966 and reissued in 1968. The original OMNITAB was designed for use with IBM 7094 computers.

At first, OMNITAB was distributed informally. In 1966 work began to redesign OMNITAB to make it as machine independent as possible and to improve it as much as possible. The responsibility for this development was first assumed by Walter Gilbert of the Computer Services Division. Two years later the responsibility for the development and maintenance of the OMNITAB computing system was assumed by David Hogben and Sally T. Peavy, respectively, of the Statistical Engineering Laboratory.

In 1969 development of OMNITAB stopped temporarily to allow time to thoroughly test and document the system. The result was the release of the "OMNITAB II Magnetic Tape and Documentation Parcel" to the National Technical Information Service for official distribution in November 1970. Less than one year later, the documentation, consisting of four publications, was completed. The version distributed by NTIS is known as OMNITAB II Version 5.0.

After 1971, development of OMNITAB resumed. Errata have been continuously sent to recipients of the "OMNITAB II Magnetic Tape and Documentation Parcel." However, improvements and additions have not been released.

In 1977, a new version of OMNITAB should be released to NTIS, which will incorporate all the many improvements made since 1971. The new version will be called OMNITAB 1977. (It will be known within the National Bureau of Standards as OMNITAB II Version 6.0) In 1970 OMNITAB became known as OMNITAB II to emphasize the machine independent feature of the computing system. For simplicity the Roman numeral II is now being replaced by the more informative 1977.

Completely new documentation will be prepared for the OMNITAB 1977 general-purpose statistical, numerical and data analysis computing system. This report is an interim document to supplement "OMNITAB II User's Reference Manual," NBS Technical Note 552. The supplement will describe all improvements and additions which have been made since 1971. No attempt is made here to correct the numerous typographical errors in NBS Technical Note 552. That will be done later when this supplement is merged with NBS Technical Note 552 in a new

OMNITAB reference manual. It is assumed that readers of this supplement are thoroughly familiar with the "OMNITAB II User's Reference Manual."

An effort was made to make OMNITAB II as machine independent as possible with great success within the state of the art in 1970. OMNITAB 1977 will be much more portable and essentially the only machine dependent features will be machine constants and the size of the system, which limits the implementation of OMNITAB to large computers (the equivalent of approximately 64K 36-bit words of memory with overlay).

OMNITAB II has been used primarily with batch processing. OMNITAB 1977 will be fully interactive with either wide or narrow input devices.

### Notation

The structure of this supplement closely parallels the structure of the "OMNITAB II User's Reference Manual," NBS Technical Note 552. PART A of NBS TN 552 is not modified. The supplement contains PART B SUPPLEMENT, PART C SUPPLEMENT and PART D SUPPLEMENT for PART B, PART C and PART D of NBS TN 552. A table of contents is given for the entire Supplement.

The numbering of sections will follow the style of NBS TN 552 with the following modifications.

- (1) A revision of a section will have an asterisk (\*) after the section number, as in

4.2\* Analysis Of One Column Of Data.

- (2) An addition to a section will have a plus (+) after the section number, as in

3.3+ Manipulative Operations.

- (3) A new section will have a period and a numeral after the number of the section which it is supposed to follow. For example, section

1.5.1 Extended Plotting Capability.

is to be inserted after

1.5 Plotting Data

and before

1.6 Optional Forms Of Readable Printing.

If more than one section is to be inserted, then the sections will be numbered consecutively as 1.5.1, 1.5.2, etc.

### Acknowledgments

OMNITAB is a very large system whose great success is the result of efforts by many people. First, there was the original design by Joseph Hilsenrath. Then came the work of the other authors of NBS Handbook 101 and the many people who contributed ideas and programming effort for that book. Next came the work of Walter Gilbert, Ruth N. Varner and many others who contributed to the development of OMNITAB II Version 5.0.

In addition to above persons, OMNITAB 1977 has been developed with the assistance of Shirley G. Bremer. Every other member of the staff of the Statistical Engineering Laboratory has contributed valuable suggestions for improvements and editorial assistance.

James J. Filliben made major contributions to the documentation and development of the probability plotting instructions, the STATISTICAL analysis instruction, and the STATPLOTS instruction. Roy H. Wampler provided a subprogram needed for the STATISTICAL analysis instruction and gave assistance during the implementation of new algorithms for the MINVERT and SOLVE instructions. Irene Stegun and Ruth Zucker contributed a program for evaluating special integrals. G. M. Furnival, Yale University, and R.W. Wilson, USDA Forest Service, provided a program used for the BESTCP instruction. Wesley L. Nicholson developed the algorithms needed for the instructions used to produce stem and leaf displays. The development of the editing instructions was stimulated by reading "Student Handbook for the Minitab Statistical Computing System," Thomas A Ryan, Brian L. Joiner and Barbara F Ryan, Pennsylvania State University (1975).

OMNITAB is a mission-oriented computing system designed to meet scientific needs. Based upon experience with scientific problems, many users, too numerous to mention, have made invaluable suggestions for improvements, additions and corrections.

The National Highway Traffic Safety Administration, Department of Transportation, provided financial support for accelerated development of OMNITAB 1977; particularly for the development of instructions for table making and data editing. The Computer Services Division, ICST, provided computing support for the entire development of OMNITAB 1977.

TABLE OF CONTENTS

Preface . . . . .	1
Notation . . . . .	2
Acknowledgments . . . . .	2
Table Of Contents . . . . .	4
<u>PART B SUPPLEMENT: THE OMNITAB II COMPUTING SYSTEM</u> . . . . .	6
1. HOW TO USE OMNITAB II . . . . .	7
1.4* NBS Operating System Control Cards . . . . .	7
1.6* NRMAX . . . . .	7
3. DIAGNOSTIC FEATURES AND ACCURACY . . . . .	8
3.2* Fatal Errors . . . . .	8
3.3* Arithmetic Faults . . . . .	9
3.4* Informative Diagnostics . . . . .	9
3.6* Accuracy Of Instructions . . . . .	11
4. FOR MORE EFFECTIVE USE OF OMNITAB II . . . . .	12
4.6.1 Use Of FASTRAND . . . . .	12
<u>PART C SUPPLEMENT: DESCRIPTIONS OF INSTRUCTIONS</u> . . . . .	16
1. ENTERING AND PRINTING DATA . . . . .	17
1.1.1 Use Of OMNITAB With Terminals . . . . .	17
BRIEF, FULL, INTERACTIVE, LENGTH, LOCAL, REMOTE, TERMINAL, WIDTH	
1.1.2 Use Of Labels . . . . .	21
ALABEL, LABEL, MLABEL	
1.1.3 Multilingual OMNITAB II . . . . .	26
DANSK, DEUTSCH, ENGLISH, ESPANOL, FRANCAIS, ITALIANO, JAPANESE,	
NEDERLANDS, NORSK, PORTUGESE, SLOVENE, SVENSKA, YUGOSLAV	
VOCABULARY	
1.3* Common Printing Instructions . . . . .	26
1.5.1 Extended Plotting Capability . . . . .	26
CPLOT, FOURPLOTS, NCPLLOT, NPLLOT, TWOPLOTS	
1.5.2 Use Of Calcomp Plotter . . . . .	35
CALCOMP AXIS, CALCOMP FAST, CALCOMP PAPER, CALCOMP PLOT,	
CALCOMP SIZE, CALCOMP SLOW, CALCOMP SPEED, CALCOMP TAPE "L"	
1.6* Optional Forms Of Readable Printing . . . . .	45
1.7* Formatted Printing And Reading . . . . .	45
1.10* Use Of Magnetic Tapes, Drums, Discs, etc. . . . .	45
BACKSPACE UNIT "L", CREAD UNIT "L", CREAD UNIT "L" "L",	
CSET UNIT "L", ENDFILE UNIT "L", READ UNIT "L", READ UNIT "L" "L",	
REWIND UNIT "L", SET UNIT "L", SKIP UNIT "L", UNIT, TAPE,	
WRITE UNIT "L", WRITE UNIT "L" "L"	
2. ARITHMETIC OPERATIONS . . . . .	49
2.6+ Data Summarization . . . . .	49
AVERAGE, DAYS, MEDIAN, PERCENTAGE, PROPORTION, RANGE, STDDEV	

3.	DATA MANIPULATION . . . . .	52
3.3+	Manipulative Operations . . . . .	52
	CENSOR EQ, CENSOR GE, CENSOR GT, CENSOR LE, CENSOR LT, CENSOR NE	
3.5.1	Editing Data . . . . .	53
	CHOOSE, CODE, DELETE, OMIT, RECODE, REPLACE, RETAIN	
4.	STATISTICAL ANALYSIS . . . . .	60
4.1*	Elementary Analysis . . . . .	60
4.1.1	Stem-and-Leaf Displays . . . . .	63
	STEM LEAF, SSTEM LEAF	
4.1.2	Statistical Plots . . . . .	72
	STATPLOTS	
4.1.3	Table Making . . . . .	78
	NTABLE AVERAGE, NTABLE FREQUENCY, NTABLE MAXIMUM, NTABLE MEDIAN, NTABLE MINIMUM, NTABLE PERCENTAGE, NTABLE PROPORTION, NTABLE RANGE, NTABLE STDDEV, NTABLE SUM, TABLE AVERAGE, TABLE FREQUENCY, TABLE MAXIMUM, TABLE MEDIAN, TABLE MINIMUM, TABLE PERCENTAGE, TABLE PROPORTION, TABLE RANGE, TABLE STDDEV, TABLE SUM	
4.2*	Analysis Of One Column Of Data . . . . .	96
4.3*	Analysis Of Groups of Data . . . . .	104
4.4*	Analysis Of Two-way Table . . . . .	116
4.4.1	Twoway Fixed Effects Design With Replication . . . . .	116
4.5*	Regression . . . . .	125
4.5.1	Selection Of Variables In Linear Regression . . . . .	139
	BESTCP	
4.6+	Correlation . . . . .	144
4.7.1	Probability Plotting . . . . .	144
	CAUCHY PLOT, DEXPONENTIAL PLOT, EXPONENTIAL PLOT, EXTREME PLOT, GAMMA PLOT, HALFNORMAL PLOT, LAMBDA PLOT, LOGISTIC PLOT, LOGNORMAL PLOT, NORMAL PLOT, PARETO PLOT, POISSON PLOT, UNIFORM PLOT, WEIBULL PLOT	
4.8+	References For Section 4 . . . . .	153
5.	NUMERICAL ANALYSIS . . . . .	154
5.1*	Special Integrals . . . . .	154
	COSINTEGRAL, EEXPINTEGRAL, EINTEGRAL, EXPINTEGRAL, HCOSINTEGRAL, HSININTEGRAL, NEGEINTEGRAL, SININTEGRAL	
5.1.1	Differences . . . . .	158
	DIFFERENCES, DIVDIFFERENCES, SDIFFERENCES, SDIVDIFFERENCES	
7.	ARRAY OPERATIONS . . . . .	161
7.4*	Properties Of An Array . . . . .	161
8.	MATRIX OPERATIONS . . . . .	162
8.5*	Matrix Analysis . . . . .	162
8.6*	Matrix Properties . . . . .	163
11.+	INDEX OF COMMANDS DESCRIBED IN PART C SUPPLEMENT . . . . .	166
	<u>PART D SUPPLEMENT: LIST OF INSTRUCTIONS</u> (arranged alphabetically) . . . . .	167

PART B SUPPLEMENT

THE OMNITAB II COMPUTING SYSTEM

## 1. HOW TO USE OMNITAB II

### 1.4\* NBS Operating System Control Cards.

Replace the first two control cards listed on page 16 of NBS TN 552 by:

```
@RUN,P/N NAMEXX,TASKNO-NAMEXX,PROJECTID,MT,MP/MC
@NBS*OMNITAB.
```

where

P = a single character F, J, M or N to denote the priority.

NAMEXX = the six character name assigned to you by the Computer Services Division.

TASKNO = the five digit task number assigned to you by the Computer Services Division.

PROJECTID = from 1 to 12 characters from the alphanumeric set A-Z and 0-9 and - and \$.  
The character \$ is not recommended.

MT = maximum run time; usually 1 for one minute. If the integer is preceded by S, the maximum time is in seconds.

MP = maximum number of pages.

/MC = maximum number of cards to be punched. Optional, use only if cards are to be punched.

### 1.6\* NRMAX.

The READ, SET and GENERATE (and all TAPE input) instructions have been modified to reset NRMAX to the number of values put in the worksheet, regardless of whether the new value is larger or smaller than the previous value.

The set of instructions

```
SET data in column 1
7, 8, 3, 6, 4
SET data in column 2
14, 11, 17
```

formerly would have left NRMAX equal to 5, but in now NRMAX is reset to 3 and the following would appear in the LIST OF DATA, INSTRUCTIONS AND DIAGNOSTICS

```
SET DATA IN COLUMN 1
7, 8, 3, 6, 4
```

```
* INFORMATIVE DIAGNOSTIC FOR THE ABOVE INSTRUCTION -
NRMAX HAS BEEN RESET FROM 0 TO 5.
```

```
SET DATA IN COLUMN 2
14, 11, 17
```

```
* INFORMATIVE DIAGNOSTIC FOR THE ABOVE INSTRUCTION -
NRMAX HAS BEEN RESET FROM 5 TO 3.
```

### 3. DIAGNOSTIC FEATURES AND ACCURACY

#### 3.2\* Fatal Errors.

Many changes have been made to the fatal error messages. Some messages have been added. Some changes were made to make the messages grammatically correct, some were made to make them technically correct, and some were made to make them clearer. A complete list of all the fatal error messages is printed below.

COMMAND DOES NOT EXIST.  
INCORRECT STATEMENT NUMBER.  
INCORRECT ARGUMENT IN INSTRUCTION.  
INSTRUCTION IS NOT ALLOWED IN REPEAT MODE.  
INSTRUCTIONS BETWEEN BEGIN AND FINISH CANNOT BE NUMBERED.  
NRMAX = 0.  
(n) IS AN INCORRECT NUMBER OF ARGUMENTS.  
COLUMN NUMBER(S) OUTSIDE (n) COLUMN WORKSHEET.  
TOO MANY NUMBERED INSTRUCTIONS.  
INSTRUCTION NUMBER NOT FOUND.  
DIMENSIONED WORKSHEET EXCEEDS LIMIT OF 12500.  
ROW NUMBER(S) OUTSIDE (n) ROW WORKSHEET.  
ARRAY OR MATRIX OUTSIDE (m) ROW X (n) COLUMN WORKSHEET.  
INTEGER ARGUMENT IS LESS THAN -8191.  
NUMBERED PERFORM INSTRUCTION EXECUTES ITSELF.  
IMPROPER TYPE OF ARGUMENT.  
INSTRUCTION MUST BE STORED.  
MATRIX IS (NEARLY) SINGULAR.  
INSUFFICIENT SCRATCH AREA.  
DEGREE IS GREATER THAN NUMBER OF NON-ZERO WEIGHTS.  
NEGATIVE WEIGHTS MAY NOT BE USED.  
INCONSISTENT ROW AND COLUMN NUMBERS.  
MISSING OR INCORRECT FORMAT OR QUALIFIER.  
NUMBER OF ARGUMENTS SHOULD BE (n).  
AN INCREMENT INSTRUCTION CANNOT INCREMENT ITSELF.  
MATRIX IS NOT SYMMETRIC.  
STORAGE COLUMN NUMBERS CANNOT EQUAL OTHER COLUMN NUMBERS.  
NRMAX IS LESS THAN 5 OR ALL MEASUREMENTS ARE THE SAME.  
VALUE OF SOME MEASUREMENT IS NOT ACCEPTABLE.  
I, J, K AND/OR L ARE NOT DEFINED CORRECTLY, OR RULE  
CANNOT DETERMINE PROPER PARAMETER VALUES FOR THESE DATA.  
THE ARGUMENTS \*NRMAX\*, \*\*NRMAX\*\*, \*V\*, OR \*\*V\*\*  
CAN ONLY BE INCREMENTED BY 0.0.  
THE PROBABILITY DISTRIBUTION PARAMETER IS NOT ALLOWED.  
LABEL OF A COLUMN (ARRAY, MATRIX) CANNOT BE A NUMBER.  
A LABEL MUST HAVE AT LEAST 1 AND LESS THAN 13 CHARACTERS.  
COLUMN (ARRAY, MATRIX) HAS BEEN PREVIOUSLY LABELED.  
NUMBER OF LABELS EXCEEDS (n) COLUMNS IN THE WORKSHEET.  
LABEL IN ALABEL OR MLABEL MUST BE FOLLOWED BY 4 NUMBERS.

### 3.3\* Arithmetic Faults.

Many changes have been made to the arithmetic messages. Some changes were made to make the messages grammatically correct, some were made to make them technically correct, some were made to improve appearance, and some were made to reduce the amount of computer storage required. A complete list of the new set of arithmetic fault messages is given below.

SORT, LOG OR RAISE OF NEGATIVE NUMBER.  
EXPONENT TOO SMALL OR TOO LARGE.  
VALUE OUT OF RANGE AND INVERSE FUNCTION CANNOT BE EVALUATED.  
X TOO LARGE FOR SIN(X) OR COS(X).  
VALUE SCALED TO AVOID OVERFLOW OR UNDERFLOW.  
DIVISION BY ZERO.  
TRIGONOMETRIC FUNCTION NOT DEFINED.  
ONE OF THE VALUES COMPARED IS ZERO. ABSOLUTE TOLERANCE USED.  
X FOR ELLIPTICAL INTEGRAL IS GREATER THAN OR EQUAL TO ONE.  
OVERFLOW FROM USE OF THE SUM ALGORITHM.  
FUNCTION NOT DEFINED FOR NEGATIVE OR ZERO VALUES.  
FUNCTION NOT DEFINED FOR ZERO VALUES.

### 3.4\* Informative Diagnostics.

Many changes have been made to the informative diagnostic messages. Some messages have been added. Some changes were made to make the messages grammatically correct, some were made to make them technically correct, and some were made to make them clearer. A complete list of all the informative diagnostic messages is given below.

TOO MUCH DATA AFTER SET, READ OR GENERATE.  
ALL DATA WERE LOST AFTER ROW (n)  
THE INSTRUCTION WAS EXECUTED, BUT CANNOT BE STORED.  
VALUE REQUESTED WAS NOT FOUND.  
COLUMN NUMBER INCORRECT OR NOT FOUND.  
THE INSTRUCTION WAS IGNORED BECAUSE ...  
ITS MEANING IS NOT CLEAR.  
A VALUE OF F LESS THAN ZERO WAS SET EQUAL TO ZERO.  
A VALUE OF DEGREES OF FREEDOM LESS THAN 1 WAS RESET TO 1.  
A VALUE OF DEGREES OF FREEDOM WAS TRUNCATED TO AN INTEGER.  
TITLE NUMBER MUST BE 1, 2, 3 OR 4 and 1 WAS USED.  
NO. ROWS NOT = TO NO. COLS. LARGEST SQUARE MATRIX WAS USED.  
AN INCORRECT ASTERISK STRING IMPLYING THROUGH WAS IGNORED.  
UNNECESSARY ARGUMENTS IN INSTRUCTION WERE IGNORED.  
MATRIX EXTENDS BEYOND (n) ROW BY (n) COLUMN WORKSHEET.  
ONLY PART OF THE MATRIX IS STORED IN THE WORKSHEET.  
INSUFFICIENT SCRATCH AREA.  
(n) = NRMAX IS NOT LARGE ENOUGH TO ALLOW ITERATION.  
1ST COLUMN OF ISETUP OR ISOLATE IS NOT MONOTONIC  
OR IS CONSTANT.  
ITERATION DID NOT FIND ANY ROOTS.  
(n) ROW WORKSHEET IS TOO SHORT TO ACCOMMODATE ALL THE VALUES  
GENERATED BY THIS INSTRUCTION.  
NO EXTREMA WERE FOUND.  
A TRIAD OF X'S WITH AT LEAST TWO IDENTICAL VALUES,  
WAS FOUND AND IGNORED.  
ONLY THE FIRST ARGUMENT IN THE INSTRUCTION WAS USED.  
FORMAT WAS NOT FOUND. READABLE FORMAT WAS USED.  
THE INSTRUCTION WAS IGNORED BECAUSE ...  
ONE, SOME OR ALL WEIGHTS ARE NEGATIVE.  
THE INSTRUCTION WAS IGNORED BECAUSE ...  
ALL WEIGHTS ARE ZERO.

THE INSTRUCTION WAS IGNORED BECAUSE ...  
 VALUE OF FUNCTION IS TOO LARGE OR TOO SMALL.  
 COLUMN NOT LONG ENOUGH TO STORE ALL NUMBERS.  
 FIRST (n) NUMBERS WERE STORED.  
 NOT ENOUGH DATA IN COLUMN TO RESTORE MATRIX/ARRAY.  
 DATA AVAILABLE WAS USED.  
 SUMS OF SQUARES DO NOT ADD UP. ABSOLUTE VALUE  
 OF (TOTAL-ROW-COL-RESIDUAL)/TOTAL EXCEEDS 5.E-7.  
 MORE THAN 50 HEAD COLUMN INSTRUCTIONS AND/OR LABELS  
 HAVE BEEN USED. ONLY THE LAST 50 HAVE BEEN RETAINED.  
 ATTEMPT TO PROMOTE FROM BELOW NRMAX = (n)  
 FIRST ARGUMENT IS RESET TO = (n)  
 ATTEMPT TO DEMOTE BELOW THE (n) ROW WORKSHEET.  
 DATA BELOW ROW (n) IS LOST.  
 X FOR ELLIPTICAL INTEGRALS IS 1.0 OR GREATER.  
 THE RESULT IS SET EQUAL TO 0.0.  
 NEGATIVE VALUE(S) WERE ENCOUNTERED BY PARTITION FUNCTION.  
 ZEROES STORED.  
 POSITIVE, INSTEAD OF NEGATIVE, TEMPERATURES USED.  
 CAUTION, USE EXPERIMENTALLY ONLY. NOT OPTIMUM IN ORDER TO  
 MAKE IT MACHINE INDEPENDENT. REFERENCES - J.B. KRUSKAL,  
 COMM. ACM,12,92. AND J.H. HALTON, SIAM REV., 12,1.  
 THE INSTRUCTION WAS IGNORED BECAUSE ...  
 COMMAND BEGINS WITH S AND STORAGE MUST BE REQUESTED.  
 NUMBER OF SIGNIFICANT DIGITS AFTER DECIMAL POINT,  
 HAS BEEN SET TO (n).  
 THE INSTRUCTION WAS IGNORED BECAUSE ...  
 ALL POINTS ARE OUTSIDE SPECIFIED LIMITS.  
 MATRIX OF PARTIAL CORRELATION COEFFICIENTS IS NOT SYMMETRIC,  
 INDICATING POSSIBLE COMPUTING ERROR.  
 PARTIAL CORRELATION COEFFICIENTS ARE NOT DEFINED.  
 (n) MEASUREMENTS MUST EXCEED (n) VARIABLES.  
 THE INSTRUCTION WAS IGNORED BECAUSE ...  
 LOWER LIMIT OF AXIS EQUALS UPPER LIMIT.  
 PRINTING OF STEM AND LEAF DISPLAY IS NOT POSSIBLE.  
 INSTRUCTION WAS TREATED AS SSTEM LEAF.  
 (n) ROWS IN WORKSHEET IS NOT ENOUGH FOR COMPLETE STORAGE.  
 DISPLAY IS (n) LINES, ONLY FIRST 99 DISPLAYED.  
 THE INSTRUCTION WAS IGNORED BECAUSE ...  
 WIDTH = (n) IS TOO SMALL OR TOO LARGE.  
 THE INSTRUCTION WAS IGNORED BECAUSE ...  
 WIDTH = (n) IS TOO SMALL FOR A PLOT.  
 WIDTH = (n) IS INSUFFICIENT FOR PAGE PLOT. BEST PLOT GIVEN.  
 THE INSTRUCTION WAS IGNORED BECAUSE ...  
 ALL NUMBERS IN THE COLUMN ARE THE SAME.  
 COMPUTING PROBLEMS ENCOUNTERED. RESULTS MAY BE MEANINGLESS.  
 THE INSTRUCTION WAS IGNORED BECAUSE ...  
 COLUMN WAS HEADED BY LABEL INSTRUCTION.  
 NRMAX HAS BEEN RESET FROM (n) TO (n).  
 IMPLIED THROUGH FOR LABELS MUST BE IN ALPHABETICAL ORDER.  
 LENGTH = (n) IS TOO LARGE AND IS RESET FOR NORMAL USE.  
 THE INSTRUCTION WAS IGNORED BECAUSE ...  
 LENGTH = (n) IS TOO SMALL FOR A PLOT.  
 ITERATION FAILED TO FIND AN EIGENVECTOR. (n) UNORDERED VALUES FOUND.

### 3.6+ Accuracy Of Instructions.

Accuracy is considered very important in the development of OMNITAB and much effort has been spent to obtain very accurate results. A special new algorithm has been used for summing numbers which is used for many instructions. Substantially better algorithms are now used for the MINVERT, MEIGEN and SOLVE instructions.

A much faster sorting procedure is now being used for the HIERARCHY, ORDER, SORT and STATISTICAL analysis instructions.

#### 4. FOR MORE EFFECTIVE USE OF OMNITAB II

##### 4.6.1 Use of FASTRAND.

This section is for users of the NBS Computer only.

A set of OMNITAB II instructions and/or data may be stored on a mass storage device (FASTRAND) for repeated use or editing. To reserve space on FASTRAND a file must be catalogued. This is done once only by inserting the following control statement after the RUN statement:

```
@ASG,UP    QUALIFIER*FILE.
```

where QUALIFIER and FILE are words with 1 to 12 alphanumeric characters.

Since a charge is made for mass storage, when it is no longer needed, the file should be decatalogued with the following statement:

```
@DELETE,C  QUALIFIER*FILE.
```

##### Entering information into FASTRAND.

Data or OMNITAB II sets of instructions can be entered as elements into a specified file. Element names consist of 1-12 alphanumeric characters.

(a) To enter a set of OMNITAB II instructions into a catalogued file, use the following:

```
@RUN...  
@ASG,A    QUALIFIER*FILE.           (catalogued file)  
@ELT,IL   QUALIFIER*FILE.ELEMENT1,,F  
          (OMNITAB II instructions)  
@END      F
```

A listing of the OMNITAB II instructions is printed along with a line identification number to the left of the instruction set.

(b) To enter a set of data into a catalogued file, use the following:

```
@RUN...  
@ASG,A    QUALIFIER*FILE.           (catalogued file)  
@ELT,IDL  QUALIFIER*FILE.DATA1,,FLG  
          (Data cards)  
@END      FLG
```

Data will be listed along with the line identification number to the left of the data.

##### Using elements on FASTRAND file.

(a) Set of OMNITAB II instructions.

A FASTRAND file can be used when a set of instructions, e.g., for generating reports, does not change from one run to another, but the data do change. Construct an element (see (a) above) consisting of all the instructions after the READ (or SET) instructions

(including a STOP instruction), but excluding the READ and OMNITAB instructions at the beginning.

A set of instructions like the following can then be used:

```
@RUN
@ASG,A    QUALIFIER*FILE.
@NBS*OMNITAB.
OMNITAB    9/29/76
READ data into columns 1***10    $ if that is where data is expected
      (data)
@ADD QUALIFIER*FILE.ELEMENT1
@FIN
```

(b) Data as an element in FASTRAND file.

Let element DATA consist of data to be analyzed with a set of OMNITAB instructions. The control cards necessary to use the element are as follows:

```
@RUN...
@ASG,A    QUALIFIER*FILE.
@NBS*OMNITAB.
OMNITAB    10/18/76
.
.
.
READ data into columns (C), (C),... (C)
@ADD      QUALIFIER*FILE.DATA
.
.
.
STOP
@FIN
```

The complete data element will be read into the worksheet and listed with the list of instructions, data and diagnostics unless a NO LIST precedes the READ command.

Editing elements within a file.

Corrections, additions and/or deletions may be made to elements containing either data or OMNITAB instructions. Whenever an element is created in a file and a listing is obtained, the successive lines (or cards) of the element are sequentially numbered to the left of the listing. In altering an element, these numbers are used on a line correction statement as follows:

-n,m

The minus sign (-), which must be in column 1, is the correction indicator. No imbedded spaces are allowed. All lines n through m in the element are to be replaced by the lines (cards) immediately following until another correction indicator is read. If no lines of correction follow the -n,m line correction statement, lines n through m are deleted.

The line correction statement -n (without a comma m) indicates that all statements after it are to be inserted into the element after line n until another correction indicator or control statement is encountered.

If corrections are to be inserted before the first line in the element, the correction lines are placed immediately after the control card

```
@ELT,UL    QUALIFIER*FILE.ELEMENT,,F
```

without specifying any insertion line number. The line correction statements must be in ascending order.

For example, suppose the element CODE is to be corrected as follows:

- i. change line number 13 from STATIS of column 4  
to  
SSTATIS of column 4,
- ii. delete the PRINT instructions with line numbers 23 and 24, and
- iii. insert a PLOT instruction after line number 5.

The set of control statements will be as follows:

```
@RUN
@ASG,A    QUALIFIER*FILE.
@ELT,UL   QUALIFIER*FILE.CODE,,F
-5
    PLOT    col 1 vs. col 2
-13, 13
    SSTATIS    col 4
-23, 24
@END      F
```

The element CODE will be updated with the corrections and a new listing with new line numbers printed. To alter an element consisting of data, the same procedure in the above example may be used to correct data lines.

#### Saving Data For Future Use.

At times part or all of the data and results in the worksheet need to be saved for use at some later date. This can be done easily by storing columns in an element on FASTRAND as follows:

```
@ASG,A    QUALIFIER*FILE.
@ASG,T    TEMP.
@USE      7., TEMP.
@NBS*QMNITAB.
QMNITAB ...
    (set of instructions)
WRITE UNIT A columns (C), (C), ..., (C)
BACKSPACE UNIT A 1
ENDFILE UNIT A
STOP
@COPY,I   TEMP., QUALIFIER*FILE.ELEMENT
```

appropriate numbers should be substituted for the columns (C), (C), ..., (C) which have to be saved. FILE denotes the name of a user's previously catalogued file and ELEMENT stands for the element where the data is to be stored.

For future use, all that needs to be done is:

```
@ASG,A    QUALIFIER*FILE.
QMNITAB
.
.
.
```

```
READ data into columns (C), (C), ..., (C)
@ADD QUALIFIER*FILE.ELEMENT
.
.
.
STOP
```

### General

The above description of the use of FASTRAND files is only applicable to the NBS UNIVAC 1108 computer under EXEC-8 control. It is not an attempt to be a complete discussion of EXEC-8 control cards. The methods described herein may be implemented in many ways using other EXEC-8 control cards. For further information check with the consultant in the computer room or the references listed below.

### References.

NBS Computer User's Guide, Technical Services Section, CSD, (1975), Section 4.

UNIVAC 1100 Operating System Programmer Reference, UP-4144 Rev. 3, Sperry Rand Corp., sections 3.7, 8.2.7, 9.5, and 18.2.

PART C SUPPLEMENT

DESCRIPTIONS OF INSTRUCTIONS

## 1. ENTERING AND PRINTING DATA

### 1.1.1 Use Of OMNITAB With Terminals.

BRIEF, FULL, INTERACTIVE, LENGTH, LOCAL, REMOTE, TERMINAL, WIDTH

The instructions described in this section were developed for interactive use of OMNITAB from terminals. The instructions BRIEF, FULL, LENGTH and WIDTH may also be used for the batch processing of instructions. All the instructions affect printing in one way or another. The instructions LOCAL and REMOTE control the location of printing. The instructions INTERACTIVE, LENGTH and WIDTH control the number of characters per line and the number of lines per page. The instructions BRIEF, FULL and TERMINAL control the amount of printing. The instructions LENGTH and WIDTH can be used very effectively to obtain plots of different sizes with either batch or remote processing.

```
:  
: BRIEF      $  no arguments  
:
```

An S before some commands, as in SFIT, suppresses all the automatic printing of the instruction. The instruction BRIEF suppresses only part of the automatic printing of all instructions which follow it. At the present time, BRIEF has no arguments and only affects the automatic printing of the STATISTICAL analysis, FIT and POLYFIT instructions.

BRIEF suppresses all pages after the first page of the automatic printing of the STATISTICAL analysis instructions. For the FIT and POLYFIT instructions, it suppresses the printing of the first page, the last two plots of the standardized residuals, the variance-covariance matrix, the second half of the analysis of variance and the fit omitting the last column (term).

BRIEF is in effect for all subsequent instructions until it is countermanded by a FULL instruction described below.

At some later time an argument may be added to the BRIEF instruction to indicate particular parts of the automatic printing which are to be suppressed.

```
:  
: FULL      $  no arguments  
:
```

A FULL instruction can be used to countermand any and all BRIEF instructions which precede it.

These two instructions make it possible to use OMNITAB II more effectively from a terminal.

```
:  
: INTERACTIVE use from a terminal  
:
```

When this instruction is used, instructions can be typed on a terminal. If an error occurs, the error message will be printed immediately after the instruction is typed, in addition to being printed in the LIST OF COMMANDS, DATA AND DIAGNOSTICS. The instruction can then be corrected and retyped. All instructions following an instruction with a fatal error will be executed, which is not the case if an INTERACTIVE instruction has not been used.

An INTERACTIVE instruction should appear at the beginning of a set of instructions and should not be stored for repeated use.

The maximum number of characters which can be printed on a line is changed from 120 to 72. See the description of WIDTH below for amplification of this effect. The line width can be modified by using the following optional form.

```
INTERACTIVE use with (c) characters per line
```

This instruction is the same as the one above; except it also allows one to specify the maximum number of characters printed on any one line. It has the same effect of using both INTERACTIVE with no arguments and the WIDTH instruction described below. If OMNITAB is used with a wide carriage terminal, the instruction:

```
INTERACTIVE 120
```

can be used to take full advantage of the wide carriage.

The argument (n) of an INTERACTIVE (n) instruction only refers to the number of characters per line to be printed. It does not affect the number of characters per line which are read. If the argument is greater than 80, only the first 80 characters (one card image) will be used by any read instruction.

```
LENGTH equal to (n) lines printed per page
```

A LENGTH instruction controls the number of lines per page and the size of plots. It only affects the printing of PRINT, PLOT, PAGE PLOT, NPLOT, CPLOT and NCLOT. (NPLOT, CPLOT and NCLOT are new instructions described in Section C1.5.1.) Its primary use is for changing the size of plots.

For any PRINT instruction, the maximum number of printed lines per page is normally 50. The number 50 does not include two lines available for the printing of the OMNITAB instruction, two lines of titles, a one-line NOTE at the bottom of the page, or blank lines between groups of numbers. The argument of a LENGTH instruction can be set to a number smaller than 50 for short pages (with batch use) or to a larger number for interactive use.

Normally, a plot has 50 intervals, or 51 divisions (lines), on the vertical axis. Smaller or larger plots can be obtained by changing the value of LENGTH. The number of intervals on the vertical axis is always some multiple of ten as indicated by the formula:

$$\text{number of intervals} = 10 \text{ times the integral part of } [(n-5)/10].$$

The following table gives a value of the argument (n) of LENGTH for different plot sizes:

<u>Intervals</u>	<u>LENGTH (n)</u>
10	15-24
20	25-34
30	35-44
40	45-54
50	55-65
60*	65-74
70*	75-84
80*	85-94
90*	95-104
100*	105-114
150*	155-164
200*	205-214
300*	305-314
400*	405-414
500*	505-514

\* These sizes are only available if an INTERACTIVE instruction has been used previously.

A LENGTH instruction can be used very effectively to produce plots of different sizes. Some examples of the use of LENGTH are given in Section C1.5.1. With a terminal, long plots can be obtained by setting the argument of a LENGTH instruction to a large number.

#### Restrictions and Informative Diagnostics

(a) The argument (n) must be positive. If it is zero, negative or not specified, the following informative diagnostic is printed:

THE INSTRUCTION WAS IGNORED BECAUSE...  
ITS MEANING IS NOT CLEAR.

(b) If the argument (n) is greater than 50 and an INTERACTIVE or TERMINAL instruction has not been used, the following informative diagnostic is printed:

LENGTH = (n) IS TOO LARGE AND IS RESET FOR NORMAL BATCH USE

Afterwards, a PRINT instruction will print a maximum of 50 lines of numbers and a PLOT instruction will have 50 intervals on the vertical axis. These are the normal conditions set by the OMNITAB control instruction.

(c) If the argument (n) of LENGTH is less than 15 and a CPLOT, NCPLOT NPLOT, PAGE PLOT, or PLOT instruction is used, the following informative diagnostic message is printed.

THE INSTRUCTION WAS IGNORED BECAUSE...  
LENGTH = (n) IS TOO SMALL FOR A PLOT.

```

:-----:
: LOCAL printing at terminal  $  no arguments  :
:-----:

```

A LOCAL instructions is used to countermand a REMOTE instruction, described below, and force all printing to be done at the terminal.

```

:-----:
: REMOTE printing on high-speed printer  $  no arguments  :
:-----:

```

When using OMNITAB from a terminal, all printing performed by OMNITAB is done at the terminal. Sometimes, a set of instructions will produce a large amount of printing, in which case it may be advantageous to use a terminal to write and correct, if necessary, a set of instructions, but have the printing done on a high-speed printer. This is done simply by using a REMOTE instruction. The effect of using a REMOTE instruction can be negated by using a LOCAL instruction.

```

:
:  TERMINAL      $  no arguments
:
:

```

A TERMINAL instruction is equivalent to the use of

INTERACTIVE 72  
BRIEF

and in addition suppresses the printing of the LIST OF INSTRUCTIONS, DATA, AND DIAGNOSTICS and the MESSAGE after the complete set of instructions has been executed.

The TERMINAL instruction is intended for interactive use with a terminal or similar peripheral device. Its use in the batch mode is possible, but unwise since the LIST OF INSTRUCTIONS, DATA, AND DIAGNOSTICS would not be printed. Caution, a TERMINAL instruction countermands any previous WIDTH instructions.

```

:
:  WIDTH set to a maximum of (c) characters per line
:
:

```

The maximum number of characters (including blanks) which is printed on any one line is (c). This instruction is normally used when instructions are typed on a terminal, but it can also be used to adjust the maximum number of columns printed on a page and the size of plots.

The OMNITAB instruction sets c = 120 for a high speed printer. An INTERACTIVE instruction, without any arguments, and TERMINAL instruction set c = 72. The INTERACTIVE instruction with the argument (c) puts WIDTH (c) into effect.

A WIDTH (c) instruction determines the maximum number of columns printed on a page by PRINT as follows:

<u>(c)</u>	<u>Number of columns</u>
15 to 29	1
30 to 44	2
45 to 59	3
60 to 74	4
75 to 89	5
90 to 104	6
105 to 119	7
120	8

Thus, if (c) is 72, up to four columns per page will be printed.

The instruction is negated by an INTERACTIVE instruction and hence should be used after the INTERACTIVE instruction which normally follows the OMNITAB instruction. It has no effect on print instructions which have a formal qualifier such as PRINT "L", ABRIDGE "L", etc., nor does it control the line width for any of the WRITE TAPE or WRITE UNIT instructions. Also, it has no effect on the printing of error messages or the LIST OF COMMANDS, DATA AND DIAGNOSTICS. Calcomp Plots are not affected by the use of a WIDTH instruction since the size of plots is determined directly by the CALCOMP instructions used.

The minimum number value of (c) allowed is 15 and the maximum value allowed is 120. If a value of (c) smaller than the allowed minimum is used or a value larger than the allowed maximum is used, the informative diagnostic:

THE INSTRUCTION WAS IGNORED BECAUSE ...  
WIDTH = (n) IS TOO SMALL OR TOO LARGE.

is printed and the number of characters per line will be 120 if the INTERACTIVE has not been used and 72 if the INTERACTIVE instruction has been used.

The size of graphs produced by PLOT instructions is changed in the horizontal direction by a WIDTH or INTERACTIVE instruction. The number of divisions of the horizontal scale is a suitable multiple of ten, plus one for an end point, so that the plot is contained in the specified width. The WIDTH instruction can thus be used very effectively to obtain plots of different sizes. A table of the number of plotting positions in the x-direction is given below.

<u>(c) of WIDTH (c)</u>	<u>Number of plotting positions, including end points</u>
59 to 68	41
69 to 78	51
79 to 88	61
89 to 98	71
99 to 108	81
109 to 118	91
119 to 120	101

Note, that if (c) is less than 59 neither PLOT nor PAGE PLOT will be executed and the following informative diagnostic will be printed:

THE INSTRUCTION WAS IGNORED BECAUSE ...  
WIDTH = (n) IS TOO SMALL OR TOO LARGE.

The size of plots produced by PAGE PLOT is unchanged if (c) is greater than or equal to 80. If it is less than 80, the above remarks for PLOT hold and if (c) is also greater than 58, the following informative diagnostic is printed:

THE INSTRUCTION WAS IGNORED BECAUSE ...  
WIDTH = (n) IS INSUFFICIENT FOR PAGE PLOT. BEST PLOT GIVEN.

### 1.1.2 Use Of Labels.

ALABEL, LABEL, MLABEL

A LABEL instruction makes it possible to use labels in instructions instead of column numbers. With the ALABEL and MLABEL instructions it is easy to use a label instead of four (or two) arguments to specify an array or matrix.

Degrees Celsius are computed from degrees Fahrenheit using the relation  $C = (F-32)/1.8$ . If the Fahrenheit degrees are in column 1, then a set of OMNITAB instructions to perform this calculation could be

```
HEAD column 1/ FAHRENHEIT
HEAD column 3/ CELSIUS
SUBTRACT 32.0 from column 1 and put in column 2
DIVIDE column 2 by 1.8 and put in column 3
PRINT columns 1 and 3
```

These five instructions are now exactly equivalent to

```

LABEL FAHRENHEIT, STORE, CELSIUS
SUBTRACT 32.0 from FAHRENHEIT and put in STORE
DIVIDE STORE by 1.8 and put in CELSIUS
PRINT FAHRENHEIT and CELSIUS

```

Column 1 is automatically used for FAHRENHEIT, column 2 for STORE, and column 3 for CELSIUS. The words FAHRENHEIT and CELSIUS are automatically used for column headings when the data are printed. It is also possible to specify the column associated with a label.

### Types of labels.

A label may have anywhere from 1 to 12 characters. A blank (space) counts as a character. The first character of a label must be either a letter or a number. If a label has only one character, that character must be a letter (from A to Z).

The characters (,) and currency symbol (\$) cannot be used in a label. A label cannot consist entirely of numbers. As indicated in the description of LABEL below, a minus sign (-) can be used to imply through as in X-Z. Hence, a minus sign should not be the second character of a label.

<u>Correct labels</u>	<u>Incorrect labels</u>
TEMPERATURE	THERMODYNAMICS
METHOD ONE	X, Y AND Z
A	+
X+1	X-1
X(1)	1976

### Use of labels in instructions.

In the ALABEL, LABEL, and MLABEL instructions, labels must be separated by commas. No other character should be used as a separator, because all the other characters can be part of a label. No descriptor words are allowed in any of these instructions. The instructions may appear anywhere in a set of instructions; except they must follow the OMNITAB and DIMENSION instructions. An ALABEL, LABEL, or MLABEL instruction cannot be stored.

When a label is used in any other instruction it must be preceded and followed by one or more blanks. Labels cannot be erased. Although a label can be reassigned to another column, array, or matrix by using a new ALABEL, LABEL, or MLABEL instruction.

In NBS Technical Note 552 commas are often used after arguments in parentheses for clarity, but they are not needed in an instruction. Below, commas are put inside parentheses to indicate that they are an essential part of the instruction.

When a LABEL (or ALABEL or MLABEL) instruction is used, it is not necessary to use the label in any instruction. The column number assigned to a label by OMNITAB either indirectly or directly (column number follows label in LABEL instruction) and the label can be used interchangeably. For example, the set of instructions:

```

LABEL TEMPERATURE, PRESSURE, YIELD
READ 1 *** 3
      (data)
MULTIPLY column 2 by 100.0 and put in col 4

```

is equivalent to the instructions:

```

LABEL TEMPERATURE, PRESSURE, YIELD
READ TEMPERATURE *** YIELD
      (data)
MULTIPLY PRESSURE by 100.0 and put in col 4

```

When many column headings are desired, one or a few LABEL instructions can be used instead of a lot of HEAD instructions.

```
ALABEL (label,) (R,) (C,) (r,) (c,), (label,) (R,) (C,) (r,) (c,), ...
```

Each label in the instruction is followed by exactly four arguments:

- (R) = the row locating the number in the upper left hand corner of the array
- (C) = the column locating the number in the upper left hand corner of the array
- (r) = the number of rows in the array
- (c) = the number of columns in the array.

These four arguments must be separated by commas; except an x may be used instead of a comma between the third and fourth arguments. A comma should not be placed at the end of the instruction.

When writing array (or matrix) operation instructions, it is common to put the word size before (r) x (c). Be sure not to do this in an ALABEL (or MLABEL) instruction, as descriptors are not allowed.

```
LABEL (label,) (label,) (label,) ...
```

Column numbers are automatically assigned to the labels. The first label is associated with column 1, the second label is associated with column 2, and so on.

Expressions such as A-Z or S-W can be used to create labels. The minus sign is used to denote through. Hence,

LABEL A-E

is equivalent to

LABEL A, B, C, D, E

Note, only a single letter is allowed. E.g., D-G is correct, but DD-GG is interpreted as a single label. The minus sign must be the second character and should not be followed by a blank space.

A LABEL instruction automatically uses labels for column headings. The label is centered. If the label has an odd number of characters, there will be one more blank on the left than on the right. The (first LABEL) instruction

LABEL METER

automatically invokes

HEAD column 1/ΔΔΔΔMETERΔΔΔ

where Δ is used to denote a blank space.

```
LABEL (label,) (C,), (label,) (C,), (label,) (C,) ...
```

In this form of LABEL, the column numbers are assigned by the user instead of being automatically assigned by OMNITAB. Any argument (C) may be deleted and the column number will be assigned by OMNITAB, using the smallest column number available.

If a minus sign is used to denote through, as in A-E, the second letter may be followed by one or more arguments, separated by commas. If a single argument is used, consecutive column numbers will be assigned to the labels, starting with the specified column number. When more than one argument is used, the arguments will be used to assign column numbers to the labels. If the number of arguments is less than the number of labels, the smallest column numbers available will be assigned to the remaining labels by OMNITAB.

In the first LABEL instruction

LABEL ALPHA, 30, S-V, 25, DATA SET 2, B-D, 10, 15, 13, 2X, J-M

column numbers will be assigned to labels as follows:

<u>label</u>	<u>column number</u>
ALPHA	30
S	25
T	26
U	27
V	28
DATA SET 2	1
B	10
C	15
D	13
2X	2
J	3
K	4
L	5
M	6

```

:-----:
: MLABEL (label,) (R,) (C,) (r,) (c,), (label,) (R,) (C,) (r,) (c,), ... :
:-----:

```

The MLABEL instruction is synonymous with the ALABEL instruction described above.

#### Diagnostics.

(1) An informative diagnostic is always printed which shows the column, array, or matrix associated with each label. For arrays and matrices, four numbers are given. Two numbers are given for the position of the number in the upper left hand corner, or pivotal point, and two numbers are printed for the size. Labels, column numbers, pivotal points and sizes of arrays and matrices are enclosed in parentheses. For example:

LABEL X, TEMP,17, METHOD, R-T

```

* INFORMATIVE DIAGNOSTIC FOR ABOVE INSTRUCTION
  (X) IS IN COL (1), (TEMP) IS IN COL (17)
  (METHOD) IS IN COL (2), (R) IS IN COL (3)
  (S) IS IN COL (4), (T) IS IN COL (5)

```

ALABEL U TIMES V, 2, 3, 4x6, X+Y, 13, 21, 8x9

```

* INFORMATIVE DIAGNOSTIC FOR ABOVE INSTRUCTION
  (U TIMES V) IS IN ROW (2) OF COL (3) SIZE (4X6)
  (X+Y) IS IN ROW (13) OF COL (21) SIZE (8X9)

```

(2) If a HEAD instruction is used for a column which is assigned to a label, the following informative diagnostic is given:

THE INSTRUCTION WAS IGNORED BECAUSE ...  
COLUMN WAS HEADED BY LABEL INSTRUCTION.

- (3) If more than 50 labels are used, the following informative diagnostic is given:

MORE THAN 50 HEAD COLUMN INSTRUCTIONS AND/OR LABELS  
HAVE BEEN USED. ONLY THE LAST 50 HAVE BEEN RETAINED.

The first labels (or headings) in are the first out.

#### Fatal errors.

- (1) The following fatal error message is printed if a label is a number:

LABEL OF A COLUMN (ARRAY, MATRIX) CANNOT BE A NUMBER.

- (2) The following fatal error message is printed if a label has more than 12 characters:

A LABEL MUST HAVE AT LEAST 1 AND LESS THAN 13 CHARACTERS

- (3) If two labels are assigned to the same column or if two arrays (matrices) are assigned the same pivotal point, the following fatal error message is printed:

COLUMN (ARRAY, MATRIX) HAS BEEN BEEN PREVIOUSLY LABELED.

No check is made to determine whether a label has been used more than once. If a label is used more than once, the last use will determine the column number or pivotal point in use.

- (4) If the number of labels exceeds the number of columns in the worksheet, the following fatal error message is printed:

NUMBER OF LABELS EXCEEDS (n) COLUMNS IN WORKSHEET.

where (n) is the number of columns in the worksheet.

- (5) If a label in ALABEL or MLABEL is followed by less than four arguments, the following fatal error message is printed:

LABEL IN ALABEL OR MLABEL MUST BE FOLLOWED BY 4 NUMBERS.

This error can occur if the commas between the arguments are accidentally omitted or if a comma is not put after any label.

#### Caution.

Instructions can now be written in a more natural form. Descriptors are still allowed. However, care should be exercised to avoid using a descriptor when a label should be used or vice versa.

Care should be exercised in using a single letter as a label, because the label might be interpreted as a format, unit, or variable qualifier. In

```
LABEL A, Y  
READ A, Y
```

the A in the second instruction is considered a qualifier, not a label. If the letter A in the READ instruction is meant to be a label, one or more descriptors should be inserted before it, as in

### 1.1.3 Multilingual OMNITAB II.

DANSK, DEUTSCH, ENGLISH, ESPANOL, FRANCAIS, ITALIANO, JAPANESE,  
NEDERLANDS, NORSK, PORTUGESE, SLOVENE, SVENSKA, YUGOSLAV, VOCABULARY

Twelve instructions were added so that instructions can be written in languages other than English. It is possible to switch from one language to another by simply inserting the appropriate command. Hence, the command ENGLISH is also given to return the system to its normal mode of operation. An additional command, VOCABULARY, prints the first six letters of each word of every command in the system for both English and the last language specified by one of these commands.

Translations for some of the languages have not been made. For other languages, the translation is tentative and subject to revision. Updating of translations will not, in general, keep up with the addition of new instructions. In all cases, translations are not given for the magnetic tape instructions or the Bessel function instructions.

### 1.3\* Common Printing Instructions.

A FIXED instruction with argument zero has been modified to print integers without the decimal point. Zero is now printed without a zero before the decimal point when FIXED (n) is used and n is greater than zero.

### 1.5.1 Extended Plotting Capability.

CPLOT, FOURPLOTS, NCLOT, NPLOT, TWOPLOTS

```
⋮  
⋮ CPLOT col (C), symbol (E)...col (C), symbol (E) versus column (C) ⋮  
⋮
```

A CPLOT instruction is similar to a PLOT instruction; except the plotting symbol is specified instead of being automatically assigned. Much greater flexibility is available in the use of plotting symbols.

The argument which specifies the plotting symbol always is placed immediately after the column number designating the variable for the vertical axis. The argument can be either a constant or a column number. If it is a constant, the specified symbol is used for each point on the curve. If the argument is a column number, the symbol specified in row one is used to plot the point in row one, the symbol specified in row two is used to plot the point in row two, and so on. This use of CPLOT can be very powerful. If the points fall into logical subsets such as for different laboratories, different test methods, etc., then a different plotting symbol can be used for each subset.

Numbers are used to indicate the required plotting symbol in accordance with the following table.

<u>Symbol</u>	<u>Constant</u>	<u>Symbol</u>	<u>Constant</u>	<u>Symbol</u>	<u>Constant</u>
0	0	G	17	W	33
1	1	H	18	X	34
2	2	I	19	Y	35
3	3	J	20	Z	36
4	4	K	21	/	37
5	5	L	22	.	38
6	6	M	23	-	39
7	7	N	24	+	40
8	8	O	25	*	41
9	9	P	26	(	42
0	10	Q	27	)	43
A	11	R	28	,	44
B	12	S	29	blank	45
C	13	T	30	=	46
D	14	U	31	\$	47
E	15	V	32	'	48
F	16				

Note, 0.0 or 10.0 can be used for the symbol zero. Whenever more than one point falls on the same plotting position, the last symbol is used.

If the number specifying a plotting symbol is less than zero or greater than 48, the CPLOT instruction will be executed as a PLOT instruction.

There are optional forms of CPLOT analogous to the optional forms of PLOT, as shown below.

CPLOT col (C) sym (E), ..., (C) (E) vertical scale from (K) to (K) vs col (C)

CPLOT col (C) sym (E), ..., (C) (E) vs col (C) horizontal scale from (K) to (K)

CPLOT (C) (E), ..., (C) (E) vertical (K) to (K) vs (C) horizontal (K) to (K)

### Examples

(a) The following set of instructions is used to illustrate the use of CPLOT, NCPLLOT described below, and LENGTH described in Section C1.1.1. The results are shown on page 29.

```

OMNITAB 2/18/77 EXAMPLE OF USE OF C PLOT AND NC PLOT
LABEL X, SINE, COSINE
GENERATE -4. (.2) 4.0 IN X
SIN OF X PUT IN SINE
COS OF X PUT IN COSINE
TITLEY SIN X COS X
TITLX GRAPH 1 2/3/77
WIDTH 75
LENGTH 30
C PLOT SINE WITH SYMBOL 11. AND COSINE WITH SYMBOL 12. VS X FROM -2.5 TO +2.5
SPACE 5
WIDTH 90
LENGTH 20
TITLEY SIN X COS X
TITLX GRAPH 2 2/3/77
NC PLOT SINE WITH SYMBOL 1. AND COSINE WITH SYMBOL 3. VS X FROM -3.0 TO +3.0
STOP

```

(b) The following table gives door temperatures (65" above floor) for five different types of fabric and four curtain conditions.

Fabric	Closed		Open	
	Light	Heavy	Light	Heavy
1	52	86	41	75
2	134	171	44	122
3	124	306	91	249
4	83	121	39	41
5	116	-	39	-

These data can be analyzed formally using an analysis of variance. On the other hand, a graphical analysis can be revealing and often more easily understood. A different symbol can be used to plot door temperature versus fabric number for each of the four curtain conditions: closed curtain of light fabric, closed curtain of heavy fabric, open curtain of light fabric and closed curtain of heavy fabric. A set of instructions, using C PLOT, for making a graphical analysis is:

```

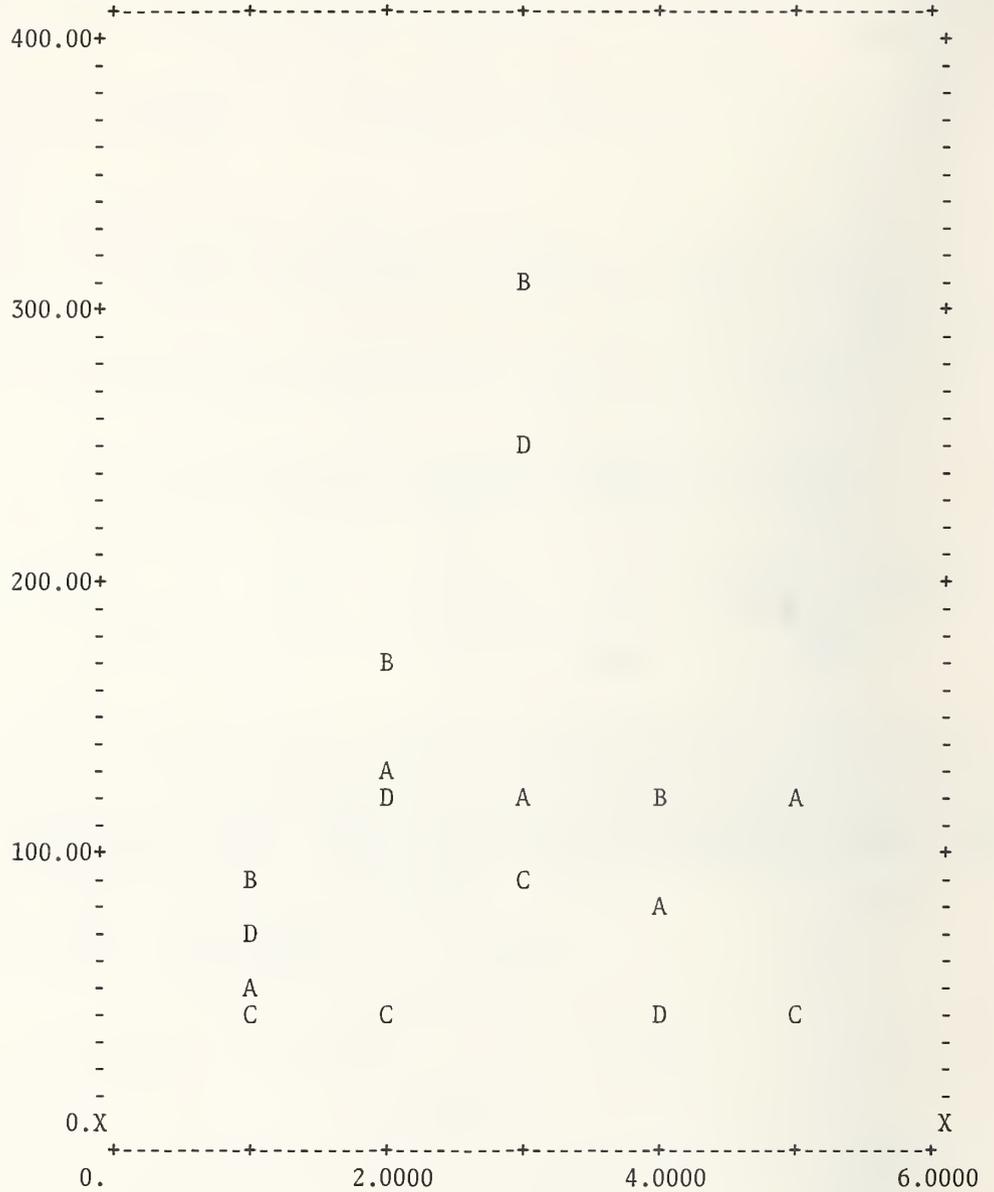
OMNITAB 2/19/77 PLOT OF FABRIC FLAMMABILITY DATA
SET DOOR TEMPERATURES IN COLUMN 1
52, 134, 124, 83, 116, 86, 171, 306, 121,
41, 44, 91, 39, 39, 75, 122, 249, 41.
$
SET CURTAIN CONDITION IN COLUMN 2
1, 1, 1, 1, 1, 2, 2, 2, 2,
3, 3, 3, 3, 3, 4, 4, 4, 4.
ADD 10.0 TO COLUMN 2 AND PUT IN COLUMN 2
$
SET TYPE OF FABRIC IN COLUMN 3
1, 2, 3, 4, 5, 1, 2, 3, 4,
1, 2, 3, 4, 5, 1, 2, 3, 4.
$
WIDTH 80
LENGTH 45
C PLOT COL 1 WITH SYMBOL IN COL 2 FROM 0.0 TO 400.0 VS COL 3 FROM 0.0 TO 6.0
STOP

```

The result is shown on page 30. Connecting like symbols by hand with colored straight lines makes the plot much easier to interpret.



ABS- COLUMN 3 ,ORD- COLUMN 1 ( ),  
 POINTS PLOTTED 18 POINTS NOT PLOTTED - OUT OF BOUND 0



```
FOURPLOTS of (C) vs (C) and (C) vs (C) and (C) vs (C) and (C) vs (C)
```

The instruction allows four plots to be printed on a single page, instead of on four separate pages. This makes a visual comparison much easier and conserves paper. Each plot has 21 by 51 divisions or 20x50 intervals. A title

PLOT OF xxxxxxxxxxxx VERSUS xxxxxxxxxxxx

appears at the top of each plot, where xxxxxxxxxxxx is either COLUMN followed by the appropriate column number or the column heading if a HEAD or LABEL instruction has been used. The number of points plotted and the number of points not plotted because they are out of bounds is printed at the bottom of each plot. The plotting symbol is a period (.). If two or more points coincide, the plotting symbol is an asterisk (\*).

If the value of WIDTH, as determined by a WIDTH, INTERACTIVE or TERMINAL instruction, is less than 120, the first two plots are printed, one below the other, on one page and the last two plots are printed, one below the other, on the next page.

The first column number, (C), in any pair is always used for the vertical axis variable and the second column number, (C), in any pair is always used for the horizontal axis variable.

```
FOURPLOTS (C),(K),(K),(C),(K),(K),(C),(K),(K),(C),(K)(K)(C),(K),(K),(C),(K),(K)
```

This form of FOURPLOTS is the same as the one above except; the upper and lower limits of the vertical and horizontal axis can be specified by a pair of constants (K), (K). Any of the eight pairs of limits, (K) (K), may be omitted or inserted. The instructions:

```
LABEL REL. INCOME, 12  
FOURPLOTS 11 vs 12, 11, 30., 40. vs 13, 11 vs 14, 140., 190., 12, 9., 11.5 vs 13, 10., 35.
```

used with the data on page 157 of NBS TN 552 produce the results shown on page 33.

```
NCPLOT (C), (E), ..., (C), (E) versus column (C)
```

A NCPLOT instruction has exactly the same relation to a CPLOT instruction as a NPRINT instruction has to a PRINT instruction. In other words, a NCPLOT instruction does not plot on a new page. Clearly, it should be used in conjunction with a LENGTH instruction. The optional forms of PLOT exist for NCPLOT as shown below.

```
NCPLOT (C),(E), ..., (C),(E) vertical scale (K) to (K) vs col (C)
```

```
NCPLOT (C),(E), ..., (C),(E) vs col (C) horizontal scale (K) to (K)
```

```
NCPLOT (C),(E), ..., (C),(E) vertical (K) to (K) vs (C) horizontal (K) to (K)
```

```
:  
: NPLOT columns (C), (C), ..., (C) versus column (C)  
:
```

A NPLOT instruction has the same relation to a PLOT instruction as a NPRINT instruction has to a PRINT instruction. A NPLOT instruction does not print the plot on a new page. It has no use unless a LENGTH instruction is used with an argument less than 50. If a LENGTH instruction is used it is possible to get more than one plot on the same page. The optional forms of PLOT exist for NPLOT. A maximum of five curves can be plotted at one time.

```
/ NPLOT cols (C), (C), ..., (C) vertical scale from (K) to (K) versus col (C)
```

```
/ NPLOT cols (C) ... (C) versus column (C) with horizontal scale from (K) to (K)
```

```
/ NPLOT cols (C) ... (C) vertical (K) to (K) vs col (C) horizontal (K) to (K)
```

```
:  
: TWOPLOTS of column (C) versus column (C) and column (C) versus column (C)  
:
```

A TWOPLOTS instruction prints two plots in the upper half of a page, instead of on two separate pages. Each plot has 21 by 51 divisions or 20 by 50 intervals. A title

PLOT OF XXXXXXXXXXXX VERSUS XXXXXXXXXXXX

appears at the top of each plot, where XXXXXXXXXXXX is either COLUMN followed by the appropriate column number or the column heading if a HEAD or LABEL instruction has been used. The number of points plotted and the number of points not plotted because they are out of bounds is printed at the bottom of each plot. The plotting symbol is a period (.). If two or more points coincide, the plotting symbol is an asterisk (\*).

If the value of WIDTH, as determined by a WIDTH, INTERACTIVE or TERMINAL instruction, is less than 120, the two plots are printed one below the other on the same page.

The first column number, (C), in either pair is always used for the vertical axis variable and the second column number, (C), in either pair is always used to specify the horizontal axis variable.

```
/ TWOPLOTS col (C) (K) to (K) vs (C) (K) to (K) and (C) (K) (K) vs (C) (K) (K)
```

This form of TWOPLOTS is the same as the one above; except the upper and lower limits of the vertical and horizontal scales can be specified by a pair of constants. Any of the four pairs of constants may be omitted or included. The instructions

```
LABEL REL. INCOME, 12  
TWOPLOTS of col 12 versus col 14 and col 13 vs col 14
```

used with the data on page 157 of NBS TN 552 produce the result shown on page 34.

PLOT OF COLUMN 11 VERSUS REL. INCOME

48.600 +	
-	
-	
-	
45.125 +	
-	
-	
-	
41.650 +	
-	
-	
-	
38.175 +	
-	
-	
34.700 +	
9.3	9.84 10.38 10.92 11.46 12.

9 POINTS PLOTTED 0 POINTS OUT OF BOUNDS NOT PLOTTED

PLOT OF COLUMN 11 VERSUS COLUMN 13

40.0 +	
-	
-	
-	
37.5 +	
-	
-	
-	
35.0 +	
-	
-	
-	
32.5 +	
-	
-	
30.0 +	
8.5	13.18 17.86 22.54 27.22 31.9

4 POINTS PLOTTED 5 POINTS OUT OF BOUNDS NOT PLOTTED

PLOT OF COLUMN 11 VERSUS COLUMN 14

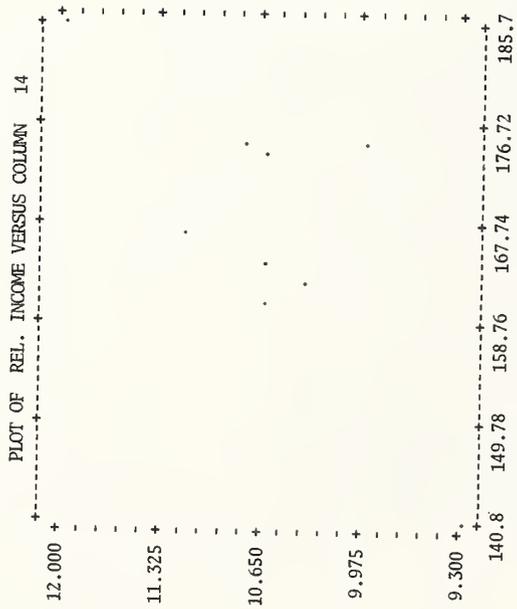
48.600 +	
-	
-	
-	
45.125 +	
-	
-	
-	
41.650 +	
-	
-	
-	
38.175 +	
-	
-	
34.700 +	
140.	150. 160. 170. 180. 190.

9 POINTS PLOTTED 0 POINTS OUT OF BOUNDS NOT PLOTTED

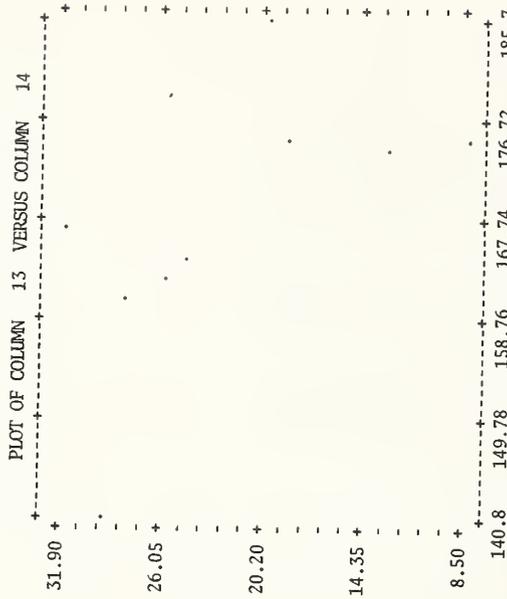
PLOT OF REL. INCOME VERSUS COLUMN 13

11.500 +	
-	
-	
-	
10.875 +	
-	
-	
-	
10.250 +	
-	
-	
-	
9.625 +	
-	
-	
9.000 +	
10.	15. 20. 25. 30. 35.

7 POINTS PLOTTED 2 POINTS OUT OF BOUNDS NOT PLOTTED



9 POINTS PLOTTED 0 POINTS OUT OF BOUNDS NOT PLOTTED



9 POINTS PLOTTED 0 POINTS OUT OF BOUNDS NOT PLOTTED

### 1.5.2 Use Of Calcomp Plotter.

CALCOMP AXIS, CALCOMP FAST, CALCOMP PAPER, CALCOMP PLOT,  
CALCOMP SIZE, CALCOMP SLOW, CALCOMP SPEED, CALCOMP TAPE

This section is only applicable if a Calcomp Digital Incremental Plotter and associated software are available.

The eight Calcomp instructions allow one to plot data using the Calcomp Digital Incremental Plotter. Since the Calcomp Plotter at NBS is not directly accessed from the NBS computer, special procedures explained below are necessary to have graphs processed. The instructions for plotting have two word commands. The first word is always CALCOMP and the second word specifies plotting conditions. The instruction CALCOMP PLOT is the basic instruction and the remaining seven instructions provide additional flexibility by taking advantage of the special features of the Calcomp Plotter.

Calcomp plots have superior quality compared to those produced by the PLOT instruction on the high-speed line printer (or terminal), but are more expensive. Unless you are sure of the form a plot should have, it would be wise to obtain one or more rough plots using the PLOT instruction before using the Calcomp Plotter.

This section describes (i) how to obtain a Calcomp plot, (ii) the basic CALCOMP PLOT instruction (with variations), (iii) the use of titles, (iv) how to change the size of the plotting paper or axes, (v) how to change the speed of the plotting pen, (vi) how to use tape units, (vii) diagnostics, and (viii) an example. Upon first reading, item (i), the first part of item (ii) and (viii) need only be read. The remaining material provides great flexibility for more sophisticated use. Additional information is given in the NBS Computer User's Guide and in "Programming for Calcomp Digital Incremental Plotters," Bulletin No. 170-E/November 1966, California Computer Products, Inc. 305 N. Muller St., Anaheim, CA 92803.

#### How To Obtain A Calcomp Plot.

Since the Calcomp Plotter at NBS is not directly accessed from the NBS computer, the information for the Calcomp Plotter is generated on magnetic tape and plotted at a later time. Four steps are required.

(1) Obtain tape reel.

Call computer desk on extension 3467 to have a tape number assigned to you. The tape may be labeled or unlabeled. The use of labeled tapes is encouraged.

(2) Control cards for tape use.

The following two control cards should appear immediately after the RUN card.

```
@MSG,W    TASKNO-NAMEXX, WRITE ENABLE TAPE yyyy  
@ASG,T    x.,8C,yyyyW
```

The symbol x stands for an integer from 7 through 12. If a CALCOMP TAPE "L" instruction is not used, set x equal to 7. Otherwise, x is 7, 8, 9, 10, 11 or 12 depending upon whether the qualifier "L" is A, B, C, D, E or F. The expression yyyy stands for the four digit magnetic tape reel number assigned to you. Note; the symbol J should follow T, if an unlabeled tape is being used. None of the TAPE or UNIT instructions should be used with a Calcomp tape.

(3) Instructions.

Write your set of OMNITAB II instructions using as many CALCOMP PLOT instructions as you wish and any of the other Calcomp instructions if

you so desire. Then send your set of instructions to the computer as usual.

(4) Send tape to Calcomp Plotter.

The instructions in (3) above do not automatically provide a plot. After you receive your output containing the LIST OF COMMANDS, DATA AND DIAGNOSTICS the tape must be sent to the Calcomp Plotter. To do this, complete Form NBS-777 as usual and in addition fill in the information for Calcomp Plots on the right, as indicated in the example on page 41. Color is normally black, but red is good for copying. Pen is usually ballpoint. Paper is either 12 or 30 inches high. (See Size Of Paper below.) Frame count is the number of graphs to be plotted. Put the reel number in the TAPE(S) box. The use of other options may be discussed at the Computer Services Division Consultant Desk.

Basic Calcomp Instruction.

The eight forms of the CALCOMP PLOT instruction generate the data to plot the graph. The different forms allow the user (i) to specify the scale of the vertical and/or horizontal axes, (ii) to plot a number of curves against a single abscissa, and (iii) to plot curves with independent abscissas (i.e. each curve defined by two arguments - an ordinate and an abscissa). A new graph is generated with each CALCOMP PLOT instruction.

The first two arguments of all forms of the CALCOMP PLOT instruction are the same. The first argument, an integer number, indicates the number of curves to be plotted for that graph. The second is either a column number or a constant with a decimal point. This argument describes the options to be used for plotting the curves. If the argument is a constant, then the option applies to all the curves on that graph. However, if the argument is a column number, then the option in row 1 of that column applies to the first curve, option in row 2 of that column to the second curve, and the option in row n to the nth curve.

Graphs are enclosed on four sides and ticked at one inch intervals. Values are printed at the tick marks along the left vertical and bottom horizontal boundaries. The numbers are in readable format. Three significant digits are used if the absolute value of the numbers is less than or equal to 100. If the absolute values exceed 100, the number of significant digits varies from 3 to 6.

```
-----:
: CALCOMP PLOT (n) curves (E) options columns (C),(C),..., (C) against (C) :
:-----:

```

Generates a graph of (n) curves with the same abscissa, defined by the last column. Both the horizontal and vertical scales are determined by OMNITAB. The maximum number of curves per graph is 97.

The general form of the option is

X.YYZZ

where X is any value 1 through 9, YY takes on values 00 through 85, and ZZ values are 00 through 99.

The value of X indicates whether symbols are to be used and how the points are to be joined (if at all) as indicated in the following table:

<u>X</u>	<u>Use Symbols</u>	<u>Join Points</u>
1	yes . . . . .	no
2	no . . . . .	yes
3	yes . . . . .	yes
4	no . . . . .	dash lines
5	yes . . . . .	dash lines
6	no . . . . .	smooth curve
7	yes . . . . .	smooth curve
8	no . . . . .	closed smooth curve
9	yes . . . . .	closed smooth curve

If X is 2 or 3, straight lines are drawn from point to point. For X=6,7,8 or 9, a curve is fitted through the points (using a modified spline fitting technique). Furthermore, the first and last point are joined for X = 8 or 9.

YY defines the particular symbol of the 85 available to be used for labeling points on the curve. If X is odd and YY is 00, then the first curve that requires a label will be labeled with the first symbol in the table below, the nth curve with the nth symbol. The first 15 symbols (01-15) are centered symbols, the rest (16-85) are lower left centered symbols. The values of YY and corresponding symbols are:

<u>Symbol</u>	<u>YY</u>	<u>Symbol</u>	<u>YY</u>	<u>Symbol</u>	<u>YY</u>
A	16	0	42	→	52
B	17	1	43	↑	53
C	18	2	44	Σ	54
D	19	3	45	>	55
E	20	4	46	<	56
F	21	5	47	—	57
G	22	6	48	∞	58
H	23	7	49	↑	59
I	24	8	50	↓	60
J	25	9	51	[	61
K	26			]	62
L	27			#	63
M	28			Δ	64
N	29	<u>Centered</u>		)	65
O	30				66
P	31	☐	01	⊖	67
Q	32	⊙	02	<	68
R	33	Δ	03	>	69
S	34	+	04	@	70
T	35	X	05	\$	71
U	36	◇	06	*	72
V	37	↑	07	(	73
W	38	X	08	%	74
X	39	Z	09	:	75
Y	40	Y	10	?	76
Z	41	☐*	11	!	77
		X	12	,	78
			13	\	79
		⊗	14	'	80
			15	;	81
				/	82
				≠	83
				≠	84
				≠	85

Every ZZth point on the curve will be labeled with the symbol requested. If ZZ is not defined (i.e., ZZ = 00) and labeling is requested, every point on the curve is labeled. Thus to label every 5th point on the curve, 05 should be the value of ZZ.

```
CALCOMP PLOT (n),(E) vertical scales (K) to (K), (C),(C),..., (C) vs (C)
```

This option is the same as the first except the user defines the range of the vertical scale (ordinate). The scale need not be increasing, i.e., the first (K) which specifies the value at the bottom of the graph can be greater than the second (K) which specifies the value at the top of the graph. The horizontal axis is determined by the instruction.

```
CALCOMP PLOT (n),(E),(C),..., (C) vs (C) horizontal scales from (K) to (K)
```

The user selects the horizontal scale (abscissa) while the instruction determines the vertical scale.

```
CALCOMP PLOT (n),(E), ordinate (K),(K),(C)...(C) vs (C) abscissa (K),(K)
```

The user selects the horizontal and vertical scales. All points which fall outside the range are not plotted.

\*\*\*\*\*

The next four forms of the CALCOMP PLOT instruction are the same as the previous forms; except every curve is defined by a pair of arguments (column numbers). For each curve that is to be plotted, an ordinate and an abscissa are given. These forms make it possible to plot curves with unrelated abscissas on the same graph. The maximum number of curves per graph is 49 if these options are used. Warning: the number of points for each curve must be the same, otherwise extraneous points will be plotted for some curves.

```
CALCOMP PLOT (n) curves, (E) options (C) vs (C), (C) vs (C),..., (C) vs (C)
```

```
CALCOMP PLOT (n),(E), vertical scale (K),(K), (C) vs (C),..., (C) vs (C)
```

```
CALCOMP PLOT (n),(E), (C) vs (C),..., (C) vs (C) horizontal scale (K),(K)
```

```
CALCOMP PLOT (n),(E) vertical (K),(K),(C) vs (C),..., (C) vs (C) horizontal (K),(K)
```

\*\*\*\*\*

### Titles.

At the top of every graph drawn by the Calcomp Plotter, the OMNITAB card is printed. On the next line, GRAPH (m) is printed, where (m) is the number printed after the CALCOMP PLOT instruction. The instructions TITLEX and TITLEY (Section C 1.5) may be used to label the horizontal and vertical axes. The information provided by TITLEY is printed on the left of the y-axis (vertical), if this instruction is used. The information supplied by TITLEX, if any, appears at the bottom of the x-axis (horizontal).

### Size Of Plotting Paper And Axes.

Chart paper roll of 120 feet in length with a plotting height (vertical axis) of 12 or 30 inches is available on the Calcomp Plotter. Twelve inch height is assumed unless otherwise specified. The standard size of the graphs is 9 inches in height by 6 inches in width for a 12 inch paper height and 27 by 18 for 30 inch paper height. Labels and headings add another  $1\frac{1}{2}$  inches to the height and  $\frac{1}{2}$  inch to the width. TITLEX and/or TITLEY add  $\frac{1}{2}$  inch more in either direction.

There are three instructions available to change the standard (i.e., assumed) height and width of the axes of a graph and the height of the plotting paper.

```
-----  
: CALCOMP AXIS (K) height of y-axis, (K) width of x-axis :  
:-----
```

The height of the y-axis and width of the x-axis are defined in inches by the two arguments. If the height and/or width is given in fraction of inches, the values will be rounded to the nearest inch.

```
-----  
: CALCOMP PAPER (n) height of paper :  
:-----
```

If (n) is not 12 or 30, an informative diagnostic is given and 12 inch height paper is assumed. This instruction should only be used once and before any CALCOMP PLOT instruction.

```
-----  
: CALCOMP SIZE (n) height of paper :  
:-----
```

The argument (n) is used to set the height of the paper to either 12 or 30 inches. This form is synonymous with CALCOMP PAPER.

```
-----  
/ CALCOMP SIZE (K) height of vertical axis (y-axis) \
```

The maximum value of (K) is the height of the paper minus 2 inches.

```
-----  
/ CALCOMP SIZE (n) height of paper (K) height of vertical axis \
```

Both the height of paper and vertical axis are defined.

```
-----  
/ CALCOMP SIZE (K) height of vertical axis, (K) length of horizontal axis \
```

The height and length of the graph are given. This form is the same as CALCOMP AXIS.

CALCOMP SIZE (n) height of paper, (K) vertical height, (K) horizontal width

The user defines the height of the paper and vertical axis and length of the horizontal axis.

#### Speed Of Plotting Pen.

The normal plotting speed is fast, but the speed at which the plotting is done can be controlled by the user using any one of the following three instructions. If one wishes to use ink pens, the plotting should be done in slow mode to permit the ink to dry before the paper or plotting arm moves. For ball point pen, fast (zip) mode is sufficient. Computer time to generate the output will not be affected by either mode. However, plotting time on the Calcomp Plotter will be longer with slow mode.

```

:-----:
: CALCOMP FAST (zip) mode      :
:-----:

```

Plotting mode is set to fast (zip) speed.

```

:-----:
: CALCOMP SLOW mode          :
:-----:

```

The plotting arm of the Calcomp Plotter will operate in slow mode to allow the ink to dry.

The following instruction has the previous two as special cases.

```

:-----:
: CALCOMP SPEED (n) $ n=0 for slow mode, n=1 for fast mode :
:-----:

```

The speed of the Calcomp Plotter is defined as either slow if n=0, or fast if n=1. The instruction

CALCOMP SPEED 1

is synonymous with the instruction

CALCOMP FAST

and the instruction

CALCOMP SPEED 0

is synonymous with the instruction

CALCOMP SLOW

#### Tape Unit.

The CALCOMP TAPE "L" instruction permits the user to direct the output for plotting to any of the logical tape units A through F. If the instruction is not used, logical tape A is assumed. NBS users must assign a tape file which may be labeled or unlabeled.

```
: _____ :  
:  CALCOMP TAPE "L" unit  :  
: _____ :
```

The graphs generated by the CALCOMP PLOT instruction are outputted on tape unit "L". The qualifier "L" is any letter A through F.

Diagnostics.

If the CALCOMP PLOT instruction is successfully executed, the following information will be printed immediately below the instruction in the LIST OF COMMANDS, DATA AND DIAGNOSTICS:

.....GRAPH m WAS PLOTTED.....

where m is an integer indicating the graph number. Because the plotting is done after the OMNITAB program is executed, this number along with the word GRAPH will be printed on the top of the graph to identify the CALCOMP PLOT instruction which generated that output for the Plotter. The user can also determine the number of graphs produced on the magnetic tape if the OMNITAB program terminates through a fatal error.

In the instructions CALCOMP AXIS and CALCOMP SIZE, the maximum height of the y-axis must be less than or equal to the height of the paper minus 2 inches. Otherwise the following informative diagnostic is given

CALCOMP PAPER WIDTH IS NOT 12 OR 30 INCHES,  
OR Y HEIGHT IS TOO LARGE.

and the axes are redefined to the standard size (9x6 or 27x18 for 12 or 30 inch paper respectively). The above diagnostic is also given and 12 inch paper assumed, if the argument of CALCOMP PAPER is not the integer 12 or 30.

Whenever the information in a TITLEY or TITLEX is too long for a graph, the following informative diagnostic is given:

TITLEX OR TITLEY TOO LONG FOR GRAPH. TITLE IS OMITTED.

The following informative diagnostic is printed if the argument in the instruction CALCOMP SPEED is incorrect:

CALCOMP PLOTTING SPEED SPECIFIED IS NOT CORRECT,  
ZIP CODE OR HIGH SPEED IS ASSUMED.

If the option(s) specified in the CALCOMP PLOT instruction is incorrect, the following informative diagnostic is given:

EITHER SYMBOL OR JOINING SPECIFIED IS INCORRECT.  
SYMBOL . OR NO JOINING IS USED.

An Example.

The following set of OMNITAB instructions produced the two graphs on pages 43 and 44 which show plots of  $\sin(x)$ ,  $\cos(x)$ ,  $\sin(x)+\cos(x)$ , and the straight line  $y = -1 + 0.5x$ .

```

OMNITAB EXAMPLES OF CALCOMP PLOT INSTRUCTION
GENERATE NOS -2.0 IN STEPS OF .5 THRU 4. IN COL 1
SIN OF COL 1 PUT IN COL 2
COS OF COL 1 PUT IN COL 3
ADD COL 2 TO COL 3 PUT IN COL 4
MULTIPLY COL 1 BY .5 THEN BY 1.0 ADD -1.0 PUT IN COL 5
GENERATE FROM 1. STEPS OF 1. THRU 4. PUT IN COL 6
RESET NRMAX TO 13
AADD 1,6 1x1 .12 PUT IN 1,6
AADD 3,6 1x1 .7102 PUT IN 3,6
TITLTY SIN (X): COS(X): SIN(X)+COS(X): Y=-1+.5*X
TITLTY FIGURE 1: OPTIONS VARY FOR EACH CURVE
CALCOMP SIZE 8. INCHES FOR Y-AXIS 6. INCHES FOR X-AXIS
CALCOMP PLOT 4 CURVES OPTIONS IN 6 Y-LIMITS (-2.,2.) FOR COLS 2 3 4 5 VS 1
TITLTY FIGURE 2: ONE OPTION FOR ALL CURVES
CALCOMP PLOT 4 CURVES OPTION 7.0103 Y LIMITS(-2.,2.) FOR COLS 2 3 4 5 VS 1
STOP
    
```

Column 6 contains the values 1.1200, 2.0000, 3.7102 and 4.0000 which are the options for curves 1, 2, 3 and 4, respectively for the first graph. The option 1.1200 requests that the first curve,  $\sin(x)$ , is to be plotted with no points joined (1), the symbol to be used is \* (12) and every point is to be labeled since the last two digits are zero. The  $\cos(x)$  curve is drawn with points joined and no labeling. The points of the third curve,  $\sin(x)+\cos(x)$ , are to be joined and every second point to be labeled with the symbol \$ (71). The straight line is drawn using dash lines.

Since the option argument in the second CALCOMP PLOT instruction is a constant (7.0103) all four curves of the second graph are to be drawn using a smooth curve which has every third point labeled with the symbol □. Note that the  $\cos(x)$  curve in graph two is smoother than the curve in graph one, because points are joined using a smooth curve rather than a straight line.

The completed form NBS-777 used to send the tape (graphs) to the Calcomp Plotter is shown below.

PR.	RUN ID	ACCOUNT NO.	TIME	PAGES	CARDS	NAME	PHONE	BIN	
		XXXXX				S. Peavy	2315		
REEL NO.	WRITE ENABLE	PERIPHERAL OPERATIONS			MAIL TO:		WILL PICK UP		
		<input type="checkbox"/> KEYPUNCH <input type="checkbox"/> VERIFY <input type="checkbox"/> INTERPRET <input type="checkbox"/> LIST CARDS _____ TIMES <input type="checkbox"/> DUPE CARDS _____ TIMES <input type="checkbox"/> DIGIDATA <input type="checkbox"/> OTHER _____ (Specify)			<input type="checkbox"/> SEND TO 2ND FLOOR <input type="checkbox"/> SEND TO 3RD FLOOR				
						SC 4020	CALCOMP		
						<input type="checkbox"/> PRINT MODE <input type="checkbox"/> PLOT MODE <input type="checkbox"/> DUPE FILM	<input type="checkbox"/> 16 MM <input type="checkbox"/> 35 MM <input type="checkbox"/> HARD COPY	INK COLOR black PAPER 12 inch	PEN ball point FRAME COUNT 2
BRKPTD REEL NO.	PGS. EST.	CDS. EST.	N O R	E X T	PART PAPER	SPECIAL FORMS	FRAME COUNT	TAPE(S)	TAPE(S)
								yyyy	
SPECIAL INSTRUCTIONS									
<input type="checkbox"/> CONT. ON REVERSE									
JOB LOADED	JOB PICKED UP		<input type="checkbox"/> OK TO RELOAD			NBS-777 (Rev. 9-73)	U.S. DEPARTMENT OF COMMERCE NATIONAL BUREAU OF STANDARDS		DATE
AT _____ BY _____	BY _____					JOB REQUEST CARD			

# GRAPH 1

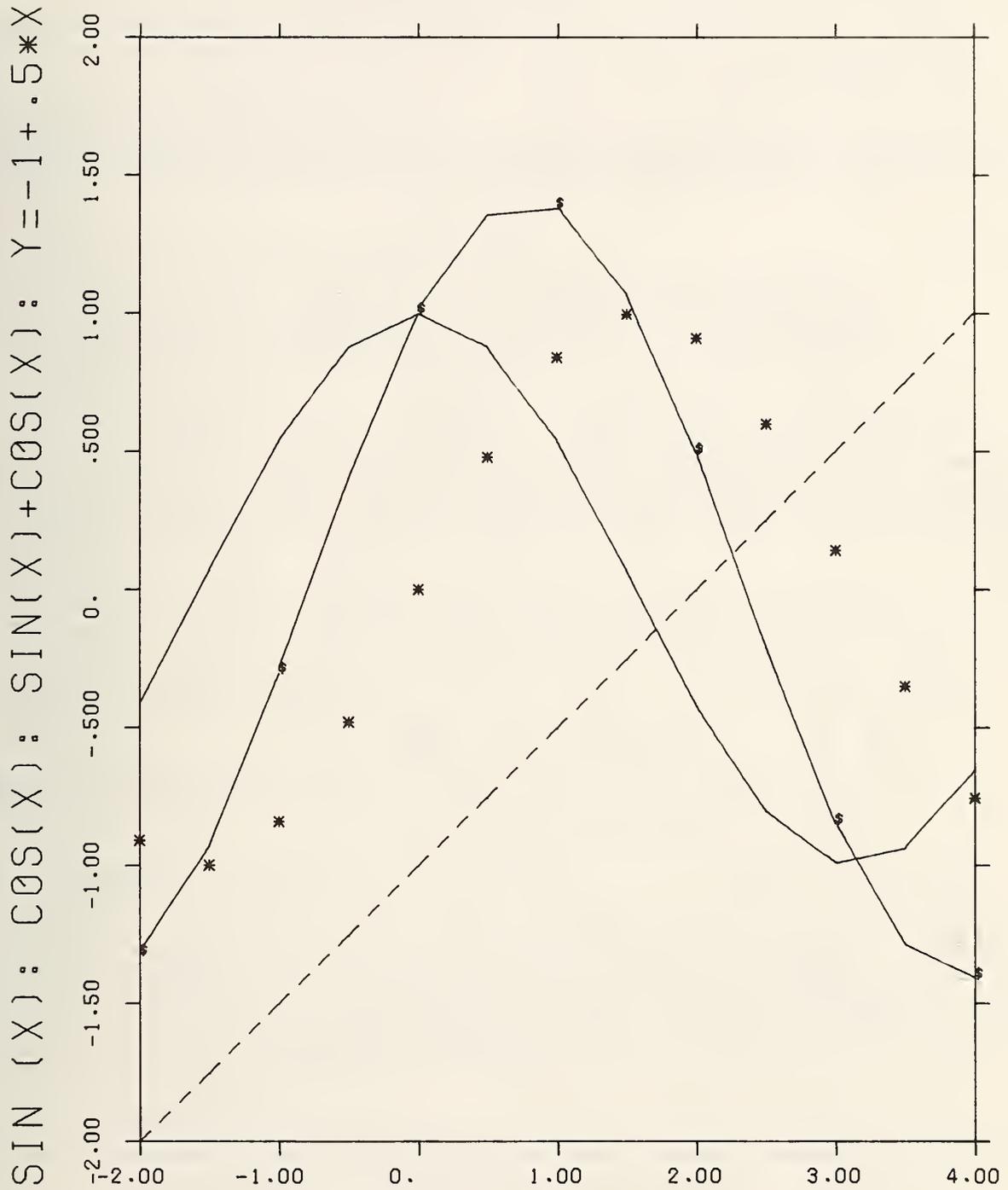


FIGURE 1: OPTIONS VARY FOR EACH CURVE

# GRAPH 2

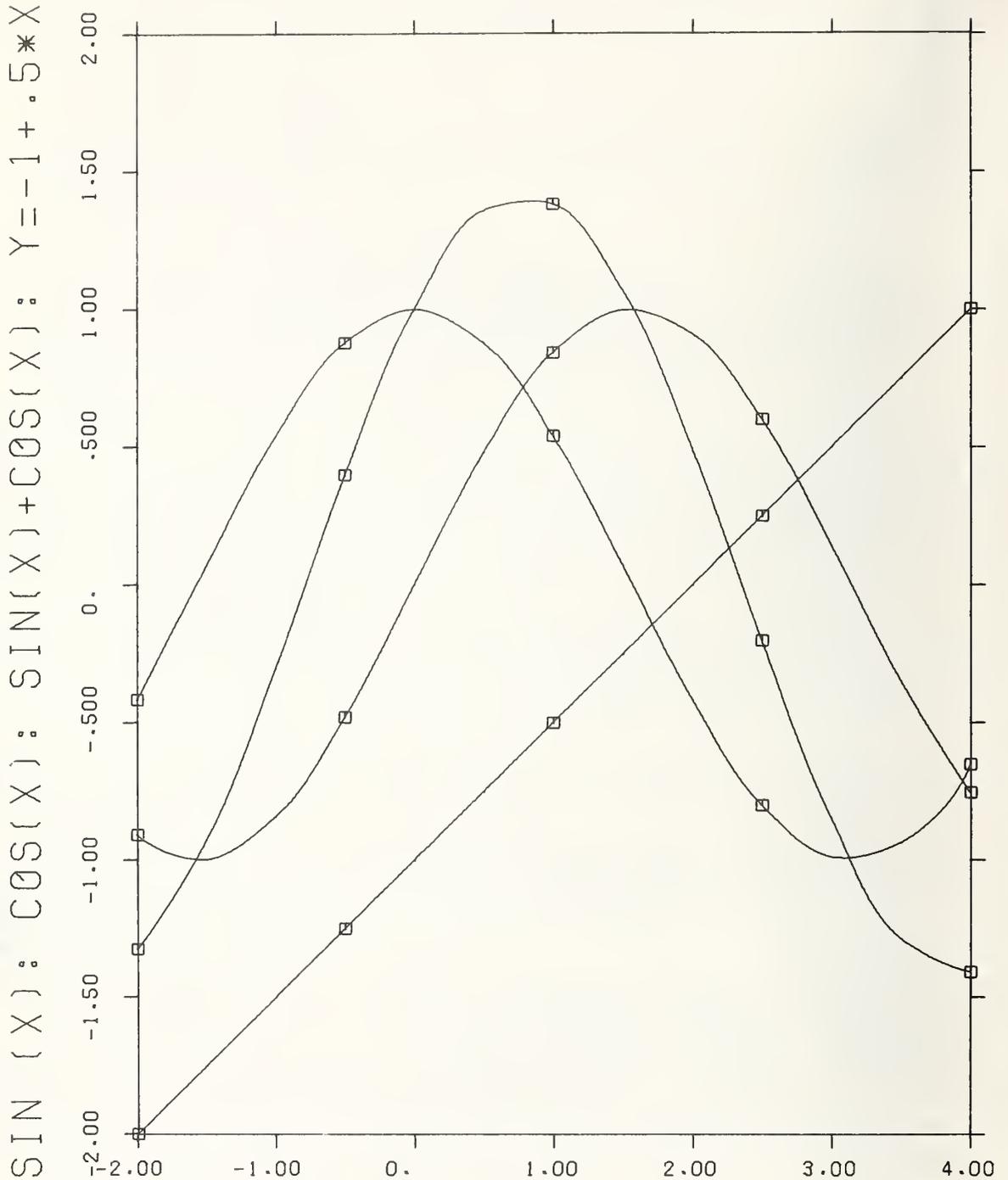


FIGURE 2: ONE OPTION FOR ALL CURVES

## 1.6\* Optional Forms Of Readable Printing.

The sentence "When numbers are printed as integers, a single blank space appears on the right." near the end of the fourth paragraph on page 73 of NBS TN 552 is no longer true. The example below should be changed accordingly.

## 1.7\* Formatted Printing And Reading.

Formatted READ with 0 count. It is now possible to read with a zero count using an An descriptor by inserting a blank card (line) at the end of the data. This is true for a READ "L" instruction, but is not valid for a READ UNIT "L" "L" instruction.

## 1.10\* Use Of Magnetic Tapes, Drums, Discs, etc.

BACKSPACE UNIT "L", CREAD UNIT "L", CREAD UNIT "L" "L", CSET UNIT "L",  
ENDFILE UNIT "L", READ UNIT "L", READ UNIT "L" "L", REWIND UNIT "L",  
SET UNIT "L", SKIP UNIT "L", UNIT, WRITE UNIT "L"  
WRITE UNIT "L" "L"

Normally, information is fed to a computer from punched cards or a card image typed on a terminal. The instructions described in this section enable one to use magnetic tapes, drums, discs, or any other medium. The average user of OMNITAB should seek additional assistance from his system programmer when working with magnetic tapes or discs. Throughout this section the word UNIT in an instruction will denote a tape, disc or drum, etc. The word TAPE can be used instead of UNIT as indicated at the end of this section under synonyms.

Each instruction in this section can be stored for repeated use. When tape operation instructions are incremented, the first qualifier should not be incremented.

Data read from tapes will not be printed with the LIST OF DATA, COMMANDS and DIAGNOSTICS. Instead, the message

(n) DATA RECORDS READ BUT NOT PRINTED

will follow the CREAD UNIT and READ UNIT instructions and diagnostic:

(n) DATA VALUES READ BUT NOT PRINTED

will follow the CSET UNIT and SET UNIT instructions.

Information on any unit or tape is grouped into logical records within blocks. Logical records may be thought of as lines on a page and blocks as pages. Unless a UNIT instruction is used, the number of characters per logical record is 80 and the number of records per block is 1. This would be the case if one were reading card images from an unblocked tape. If blocked tapes are used or if more than 80 characters per record are to be read, a UNIT instruction must be used to indicate the number of characters per logical record and, if necessary, the number of logical records per block.

## System Control Statements

This material is for the benefit of users of the NBS computer. Other users must obtain comparable information from their computer center.

When using UNIT instructions, @MSG, @ASG and @USE system control statements must be used for each medium (tape, unit, etc.) assignment. The control cards should come right after the @RUN statement and have the following form:

```

@MSG,W      TASKNO-NAMEXX, WRITE ENABLE TAPE yyyy
@ASG[,options] LABEL.,xx,yyyyW
@USE      N.,LABEL

```

where yyyy is the tape reel number and xx = 8C if the physical unit is a tape and xx= F if the physical unit is Fastrand. The MSG control statement and the letter W after the reel number are only needed if data is to be written on tape. The variable N will be 7, 8, 9, 10, 11, or 12, depending whether the qualifier of the UNIT instruction is A, B, C, D, E or F. For a complete description of the other variables and options on the control statements refer to sec. 3.7.1.1 (for ASG Fastrand medium), sec. 3.7.1.2 (for ASG tape medium) and sec. 3.7.5 (for USE) of UNIVAC 1100 Series Operating System Programmer Reference (UP-4144 Rev. 3).

### Instructions.

```

:-----:
: BACKSPACE UNIT "L" (r) logical records :
:-----:

```

This instruction backspaces unit "L" by (r) logical records. If a UNIT instruction has not been used, the instruction backspaces (r) blocks. However, if a UNIT instruction was used, the number of blocks that will be backspaced is:

$$r/r' + m,$$

where r' is the number of logical records per block and:

$$\begin{aligned}
m &= 1, && \text{if } r/r' \text{ has a remainder} \\
m &= 0, && \text{if } r/r' \text{ has no remainder.}
\end{aligned}$$

The unit is always positioned at the beginning of a block.

```

:-----:
: CREAD UNIT "L", using (r) logical records, into columns (C), (C), ... (C) :
:-----:

```

This instruction is similar to the READ UNIT "L" command; except (r) (r > 0) specifies the number of logical records to be read instead of terminating input with a logical record of zeros as in READ UNIT "L". The data within a block, which is not read, is lost.

```

:-----:
: CREAD UNIT "L" "L" format, using (n) blocks, into cols (C) ... (C) :
:-----:

```

Similar to the above instruction; except a FORMAT "L" is referenced and similar to the READ UNIT "L" "L" instructions; except (n) (n > 0) blocks of data are read.

```

:-----:
: CSET UNIT "L", using (r) logical records, into column (C) :
:-----:

```

This instruction is similar to the SET UNIT instruction (see below). Here, the indicated number of logical records (r) (r > 0) is read, whereas the SET UNIT command reads data until a logical record of zeros is encountered. If all the data from a block is not read, the rest of the information will be lost.

```

//-----//
/ CSET UNIT "L", using (r) logical records, into row (R) of column (C) /
//-----//

```

Similar to above instruction; except data is entered into row (R) and below.

```
:  
: ENDFILE UNIT "L"  
:
```

This instruction writes an end-of-file mark on the specified unit, "L". It is necessary to ENDFILE a unit after all the data has been written on that unit. Otherwise, when the data is read at some future time, extraneous information may be entered into the worksheet.

```
:  
: READ UNIT "L" into columns (C), (C) ... (C)  
:
```

This instruction is similar to the READ command; except data is read from the specified unit rather than from cards. Reading of data continues until a logical record of zeros is encountered (not a blank logical record). The execution of this instruction always starts at the beginning of a new block. Data for any particular row of all the columns specified must be completely contained within a logical record. If all the data from a block is not used, it is lost. Subsequent READ UNIT or CREAD UNIT instructions start with a new block.

```
:  
: READ UNIT "L" "L" format into columns (C), (C), ... (C)  
:
```

Same as above; except unit is read using the specified format. Reading of data is under format control and not under the control of OMNITAB. Data will be entered into the specified columns until some row contains all zeros (i.e., a block of zeros). This instruction always starts at the beginning of a new block for each row of data entered. Data for any particular row may be in one or more blocks.

Warning: One must be aware of exactly how logical records are blocked and must set up the format accordingly. Generally information should be read with the same format as it was written.

```
:  
: REWIND UNIT "L"  
:
```

The unit specified by "L" is positioned to the beginning of the unit.

```
:  
: SET UNIT "L" into column (C)  
:
```

This instruction is similar to a SET instruction; except data is read from the specified unit rather than cards. Reading of data continues until a logical record of zeros is encountered; not a blank logical record. Extra data, if any, in the block will be lost.

```
\/  
SET UNIT "L" starting with row (R) of column (C)  
\/
```

Same as above; except first datum is entered into row (R) instead of row 1.

```
:  
: SKIP UNIT "L", forward (r) logical records  
:
```

The specified unit "L" is moved forward (r) logical records. The instruction is just like a BACKSPACE UNIT instruction; except the unit is moved forward instead of backwards. The unit is positioned at the beginning of a block.

```
UNIT with (c) characters per logical record
```

The number of characters per logical record is set equal to (c). The number of logical records per block is assumed to be one.

```
UNIT with (c) characters/logical record and (r) logical records/block
```

The number of characters per logical record is set equal to (c) and the number of logical records per block is set equal to (r). This instruction should be used with blocked tapes or data. The instruction controls the number of characters per logical record and number of logical records per block for all qualifiers. Thus, if two mediums are used with differing values of (c) and (r), a different UNIT instruction must be used before the use of each medium.

Theoretically,  $c \times r$  should be less than or equal to the size of the worksheet, 12500. However, since FORTRAN input/output procedures are used, a systems programmer should be consulted to determine how large  $c \times r$  can be. For NBS users  $c \times r \leq 132$  and no check is made to see that this number is not exceeded.

```
WRITE UNIT "L" from columns (C), (C) ... (C)
```

A WRITE UNIT instruction is to tapes, discs or drums what PUNCH is to cards. Data is written on the unit "L" with c characters per logical record with r logical records per block. See the description of UNIT instruction for a definition of c and r. The default values of c and r are 80 and 1 respectively. Each row of data starts a new logical record. Let  $i =$  the number of columns specified in the instruction and  $j =$  integral part of  $c/15$ . Then the number of logical records written for each row is the integral part of  $i/j + k$ , where  $k = 0$ , if  $i/j$  has no remainder and  $k = 1$ , if  $i/j$  remainder is not zero. Each row starts a new logical record. If the last logical record contains less than c characters, it will be packed with blank characters. A row of zeros is written after all the rows of data.

If the last block has only m logical records ( $m < r$ ), then the last  $r - m$  logical records will contain zeros. The instructions FIXED and FLOATING have no effect.

```
WRITE UNIT "L" "L" format, from columns (C), (C) ... (C)
```

Same as above; except blocks are written according to the referenced format. The output from this instruction is under format control and is not controlled by the OMNITAB II computing system. Each slash (/) in the format specification or each scan of the format from the beginning will generate a new block. Furthermore, a row of zeros will be written under format control after all the data is written.

Synonyms

BACKSPACE TAPE "L", CREAD TAPE "L", CREAD TAPE "L" "L", CSET TAPE "L",  
ENDFILE TAPE "L", READ TAPE "L", READ TAPE "L" "L", REWIND TAPE "L",  
SET TAPE "L", SKIP TAPE "L", WRITE TAPE "L", WRITE TAPE "L" "L"

## 2. ARITHMETIC OPERATIONS.

### 2.6+ Data Summarization.

AVERAGE, DAYS, MEDIAN, PERCENTAGES, PROPORTIONS, RANGE, STDDEV

AVERAGE of numbers in columns (C),(C),..., (C) put in cols (C),(C),..., (C)

The optional form of AVERAGE is a generalization of the form described on the top of page 94 of NBS TN 552 which makes it easy to compute the average of more than one column.

DAYS for month (E) day (E) and year (E) put in column (C)

The instruction DAYS converts a date like 8/5/76 into a unique number. For individual dates, this number is of minor importance, unless one is coding data for a statistical analysis. However, it can be very useful for finding the number of days between any two dates. The instruction automatically adjusts for the differing numbers of days in the months and years. The definition of a leap year is somewhat complicated and not widely known, but the DAYS instruction uses it appropriately.

For the years in the 20th century (1900 to 1999), it is not necessary to include the first two digits of the year. Thus, 76 may be used instead of 1976.

To compute the number of days between January 17, 1976 and May 8, 1976, inclusive, we could use the instructions

```
READ date into columns 1, 2 and 3
  1/17/76
  5/ 8/76
DAYS for columns 1,2, and 3 put in col 4
PROMOTE 1 row column 4 into col 5
SUBTRACT col 4 from col 5 and put in col 6
ADD 1.0 to column 6 and put in column 6
ABRIDGE row 1 of col 6
```

or

```
READ date into columns 1, 2 and 3
  1/16/76
  5/ 8/76
DAYS for columns 1,2, and 3 put in col 4
PROMOTE 1 row column 4 into col 5
SUBTRACT col 4 from col 5 and put in col 6
ABRIDGE row 1 of col 6
```

to get the answer 112.

For reference purposes, the number computed for 1/1/100 is 1 and the number computed for 1/1/77 is 685,412. No month, day or year can be negative or zero. Because of the

convention described above, of being able to use only the last two digits for years between 1900 and 1999, inclusive, years between 1 AD and 99 AD cannot be used. Note, according to American, but not European, tradition, the month comes before the day rather than after it.

```
:  
: MEDIAN of numbers in column (C) put in column (C)  
:
```

If the numbers 3.1, 17.6, 11.6, 9.2 and 8.3 are in column 17, the instruction

MEDIAN of column 17 put in column 18

would put the median 9.2 into every row of column 18. The median is defined on page 120 of NBS TN 552.

```
/ MEDIAN of columns (C),(C),..., (C) put in column (C),(C),..., (C) \
```

The optional form of the MEDIAN instruction makes it possible to compute the median of several columns with a single instruction.

```
:  
: PERCENTAGES of numbers in column (C) put in column (C)  
:
```

Each number in the first named column is divided by the sum of the numbers, multiplied by 100.0 and put into the second column. If the numbers 8, 11, 14, 5 and 1 are in column 8, the instruction

PERCENTAGES of column 8 put in column 23

would put the numbers 20.512820, 28.205128, 35.897435, 12.820513 and 2.5641026 in column 23.

```
/ PERCENTAGES of columns (C),(C),..., (C) put in column (C),(C),..., (C) \
```

The optional form of the PERCENTAGES instruction makes it possible to compute the percentages of numbers in several columns.

```
:  
: PROPORTIONS of numbers in column (C) put in column (C)  
:
```

A PROPORTIONS instruction is similar to the PERCENTAGES instruction, described above; except the numbers in the second column are not multiplied by 100. If the numbers 8, 11, 14, 5 and 1 are in column 36, the instruction

PROPORTIONS of column 36 put in column 52

would put the numbers 0.2051282, 0.28205128, 0.35897436, 0.12820513 and 0.025641026 in column 52.

PROPORTIONS of columns (C),(C),..., (C) put in columns (C),(C),..., (C)

The optional form of the PROPORTIONS instruction makes it possible to compute the proportions of numbers in several columns.

RANGE of numbers in column (C) put in column (C)

The difference between largest and smallest numbers in the first column is put in the second column. If the numbers 3.1, 17.6, 9.2, 11.6 and 8.3 are in column 31, the instruction

RANGE of column 31 put in column 32

would put the number 14.5 in every row of column 32.

RANGE of columns (C),(C),..., (C) put in columns (C),(C),..., (C)

The optional form of the RANGE instruction makes it possible to compute the range for several columns of data.

STDDEV of numbers in column (C) put in column (C)

Compute the standard deviation, as defined on page 120 of NBS TN 552, of the numbers in the first column. If the numbers (see Section B3.5 of NBS TN 552) 9000, 9001 and 9003 are in column 6, the instruction

STDDEV of column 6 put in column 7

would put the number 1.5275252 in every row of column 7.

STDDEV of columns (C),(C),..., (C) put in columns (C),(C),..., (C)

The optional form of the STDDEV instruction makes it possible to compute the standard deviations of numbers in several columns.

### 3. DATA MANIPULATION.

#### 3.3+ Manipulative Operations.

CENSOR EQ,  
CENSOR LE,

CENSOR GE,  
CENSOR LT,

CENSOR GT,  
CENSOR NE

The following six instructions are modifications of the CENSOR instruction with no modifier described on page 102 of "OMNITAB II User's Reference Manual."

CENSOR EQ col (C) for values equal to (E), replace by (E), put in col (C)

CENSOR GE col (C) for nos greater than or eq. to (E), replace by (E), put in (C)

CENSOR GT col (C) for values greater than (E), replace by (E), put in col (C)

CENSOR LE col (C) for nos less than or equal to (E), replace by (E), put in (C)

CENSOR LT col (C) for values less than (E), replace by (E), put in col (C)

CENSOR NE col (C) for values not equal to (E), replace by (E), put in col (C)

A CENSOR LE instruction is synonymous with a CENSOR instruction with no modifier.

For the numbers

column 11: 2.3, 7.6, 5.2, 8.3, 4.2

column 12: 3.4, 5.8, 5.2, 4.7, 6.3

the instructions

CENSOR EQ col 11 for 7.6, replace by 0.0, put in col 21  
CENSOR GE col 11 for col 12, replace by -1.0, put in col 22  
CENSOR GT col 11 for 4.5, replace by col 12, put in col 23  
CENSOR LE col 11 for col 12, replace by -1.0, put in col 24  
CENSOR LT col 11 for 6.0, replace by col 12, put in col 25

CENSOR NE col 11 for col 12, replace by -1.0, put in col 26

would leave columns 11 and 12 unchanged and put the following numbers in columns 21 to 26:

<u>21</u>	<u>22</u>	<u>23</u>	<u>24</u>	<u>25</u>	<u>26</u>
2.3	2.3	2.3	-1.0	3.4	-1.0
0.0	-1.0	5.8	7.6	7.6	-1.0
5.2	-1.0	5.2	-1.0	5.2	5.2
8.3	-1.0	4.7	8.3	8.3	-1.0
4.2	4.2	4.2	-1.0	6.3	-1.0

### 3.5.1 Editing Data.

CHOOSE, CODE, DELETE, OMIT,  
RECODE, REPLACE, RETAIN

The instructions described in this section are useful for editing data or for data manipulation in the following problem areas.

1. Deletion. Certain data are to be deleted for any one of a number of reasons:  
(a) coded missing observations, (b) outliers, (c) possible outliers, (d) analysis of subsets of data, etc.
2. Retention. This is the complement of deletion above and is done for similar reasons. A person's height equal to -10cm should be deleted as an outlier. At the same time, we would want to retain all positive heights, knowing that all negative heights are outliers by definition.
3. Coding. There are a number of reasons for coding data. For example, to  
(a) specify subsets for further analysis, (b) simplify data, (c) aggregate values to be treated alike, e.g., all persons aged 10 to 14, (d) identify missing observations, etc.

Missing observations have to be identified. A blank is unsatisfactory. Sometimes zero can be used, but at other times it cannot be used. Missing observations must be coded and subsequently deleted from further analysis. The coded value must be outside the known range of permissible values.

The instructions OMIT and DELETE are useful for deleting data, CHOOSE and RETAIN for retaining data, and CODE, RECODE, and REPLACE for coding data. DELETE is similar to OMIT; except it can operate on several columns simultaneously. Also, RETAIN is similar to CHOOSE; except it can operate on several columns simultaneously.

The CHOOSE, DELETE, OMIT, and RETAIN instructions all reset NRMAX. The following informative diagnostic message (see Section B1.6) is always printed:

NRMAX HAS BEEN RESET FROM (m) TO (n)

where (m) is the value of NRMAX before the instruction is executed and (n) is the value of NRMAX after the instruction is executed.

Caution. Because of rounding, truncation, or conversion errors, a number in the computer may be slightly different from what we think it should be. For example, a number believed to be 8.5 may actually be something like 8.4999999 or 8.5000001. If an instruction is looking for numbers equal to 8.5, it may not succeed when it should. This is not a

serious problem. However, when it does (or might) occur, a ROUND (n), (C), (C) instruction can be used to avoid any difficulty.

```
.....:
: CHOOSE rows with (K) in column (C) and put in column (C) :
:.....:
```

All numbers in the first column equal to (K) are stored in the second column. Zeros are put at the bottom of the second column. NRMAX is reset to the number of rows in the first column which have the value (K).

```
...../
/ CHOOSE rows with numbers between (K) and (K) in column (C) and put in col (C) /
...../
```

All numbers in the first column between (K) and (K), inclusive, are stored in the second column. Zeros are put at the bottom of the second column and NRMAX is reset accordingly.

```
...../
/ CHOOSE rows with (K) in (C), corresp. rows of (C),..., (C) put in (C),..., (C) /
...../
```

The number of arguments in this form of CHOOSE is  $1 + 2n$ , where  $n = 1$ , or  $2$ , or  $3$ , ..., up to  $49$ . There are two sets of  $n$  column numbers. If (K) is found in any row of the first column of the first set of columns, that row of every column in the first set is included in the second set of columns. The other rows are omitted. Zeros are put at the bottom of the second set of columns and NRMAX is reset.

```
...../
/ CHOOSE from (K) to (K) in (C), corresp. rows of (C),..., (C) put in (C),..., (C) /
...../
```

This fourth form of CHOOSE bears the same relation to the second form as the third form does to the first. The instruction has  $2 + 2n$  arguments, where  $n$  can be any number from  $1$  to  $48$ . For the data in columns 1 to 4 shown below, the instructions

```
CHOOSE rows with 6.0 in column 1 and put in column 5
RESET 15
CHOOSE rows with numbers between 5.0 and 8.0 in column 1 and put in column 6
RESET 15
CHOOSE rows with 6.0 in col 1 corr row of col 2 put in cols 7 and 8
RESET 15
CHOOSE rows between 5.0 and 8.0 in col 1 corr rows of col 2 put in cols 9 and 10
```

would put data in columns 5 to 10 as shown in the following table:

Row/Column	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>8</u>	<u>9</u>	<u>10</u>
1	6	25	40	101	6	6	6	25	6	25
2	7	23	48	-1	6	7	6	25	7	23
3	7	25	40	104	6	7	6	-3	7	25
4	6	25	49	109	6	6	6	25	6	25
5	-2	26	40	106	6	7	6	23	7	24
6	7	24	48	102	6	6	6	25	6	-3
7	6	-3	40	103		6			6	25
8	6	25	49	108		6			6	23
9	6	23	-2	105		7			7	24
10	7	24	48	108		6			6	25
11	-1	24	48	102		7			7	26
12	6	25	49	101		7			7	25
13	7	26	48	-1		7			7	26
14	7	25	48	103		0			0	0
15	7	26	49	107		0			0	0

For the first CHOOSE instruction, the number 6 occurs in the six rows 1, 4, 7, 8, 9, and 12 of column 1. The numbers in the other rows of column 1 are not put into column 5. NRMAX is reset to  $6 = 15 - 9$ . For the second CHOOSE instruction, all the numbers in column 1 are between 5 and 8; except the numbers in rows 5 and 11. The numbers in rows 5 and 11 of column 1 are not put into column 6. NRMAX is reset to  $13 = 15 - 2$ . The third CHOOSE instruction behaves just like the first and in addition for each row that has a six in column 1 the corresponding row of column 2 is put into column 8. NRMAX is reset to  $6 = 15 - 9$ . The fourth CHOOSE instruction behaves just like the second, but whenever a row from column 1 is put into column 9 the same row of column 2 is put into column 10.

```

:-----:
: CODE column (C) using length (K) put in column (C) :
:-----:

```

The CODE instruction divides the numbers in the first column into intervals of length (K); starting with the smallest number in the column. All numbers in the first interval are assigned the value 1, all numbers in the second interval are assigned the value 2, etc. and the results are put in the column designated by the third argument. The instruction is useful for aggregating values to be treated alike. For example, all persons aged 10 to 14 could be put into one class. If the numbers 12.6, 36.4, 27.1, 6.2 and 3.7 are in column 1, the instruction

CODE column 1 using length 5.0 put in column 4

would put the numbers 2, 7, 5, 1, and 1 into column 4.

```

/-----\
/ CODE column (C) starting at (K) using length (K) put in column (C) \
\-----/

```

This form of the CODE instruction is the same as the one above; except the starting value, indicated by the second argument, is specified instead of being assigned automatically the minimum value in the column. All numbers in the first column which are less than the specified starting value are assigned the coded value 0.0.

For the numbers 12.6, 36.4, 27.1, 6.2, and 3.7 in column 1, the instruction

CODE column 1 starting at 0.0 using length 5.0 put in column 2

would put the numbers 3, 8, 6, 2, and 1 into column 2. On the other hand, the instruction

CODE column 1, starting at 10.0, using length 5.0, put in column 3

would put the numbers 1, 6, 4, 0, and 0 into column 3. In the last instruction, the starting value 10.0 exceeds the minimum value 3.7. The values 6.2 and 3.7, which are below the starting value, are coded 0.0.

```

:-----:
: DELETE rows having (K) in rows of any cols (C),..., (C) put in cols (C),..., (C) :
:-----:

```

If the number (K) is in a row of any column in the first set of columns, that row is not put into the second set of columns. The other rows in the first set of columns are retained in the second set of columns. Zeros are put at the bottom of the second set of columns and NRMAX is reset to the number of rows retained. The instruction

DELETE all rows having -1.0 in any of cols 1,2,3, and 4 and put in cols 11,12,13, and 14 would produce the following results for the numbers shown in columns 1 to 4:

Row/Column	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>11</u>	<u>12</u>	<u>13</u>	<u>14</u>
1	6	25	40	101	6	25	40	101
2	7	23	48	-1	7	25	40	104
3	7	25	40	104	6	25	49	109
4	6	25	49	109	-2	26	40	106
5	-2	26	40	106	7	24	48	102
6	7	24	48	102	6	-3	40	103
7	6	-3	40	103	6	25	49	108
8	6	25	49	108	6	23	-2	105
9	6	23	-2	105	7	24	48	108
10	7	24	48	108	6	25	49	101
11	-1	24	48	102	7	25	48	103
12	6	25	49	101	7	26	49	107
13	7	26	48	-1				
14	7	25	48	103				
15	7	26	49	107				

The number -1 occurs in row 2 of column 4, row 11 of column 1, and row 13 of column 4. Hence, rows 2, 11, and 13 are deleted when the results are stored in columns 11 to 14. NRMAX is reset to 15 - 3 = 12.

```

//-----//
// DELETE rows with numbers between (K) and (K) in (C),..., (C) put in (C),..., (C) //
//-----//

```

All rows in the first set of columns with a number between (K) and (K), inclusive, are not included in the second set of columns. The other rows in the first set of columns are retained in the second set of columns. Zeros are put at the bottom of the second set of columns and NRMAX is reset to the number of rows retained. The instruction

DELETE all numbers between -10.0 and 0.0 in cols 1,2,3, and 4 put in cols 15,16,17, and 18 would produce the following results for the numbers shown in columns 1 to 4:

Row/Column	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>15</u>	<u>16</u>	<u>17</u>	<u>18</u>
1	6	25	40	101	6	25	40	101
2	7	23	48	-1	7	25	40	104
3	7	25	40	104	6	25	49	109
4	6	25	49	109	7	24	48	102
5	-2	26	40	106	6	25	49	108
6	7	24	48	102	7	24	48	108
7	6	-3	40	103	6	25	49	101
8	6	25	49	108	7	25	48	103
9	6	23	-2	105	7	26	49	107
10	7	24	48	108				
11	-1	24	48	102				
12	6	25	49	101				
13	7	26	48	-1				
14	7	25	48	103				
15	7	26	49	107				

In this example, the six rows 2, 5, 7, 9, 11, and 13 contain numbers in columns 1 to 4 which are in the range -10.0 to 0.0 and are deleted. All the numbers in the remaining rows of columns 1 to 4 are outside the range -10.0 to 0.0 and these rows are retained. NRMAX is reset to  $15 - 6 = 9$ .

```

:-----:
: OMIT rows with (K) in column (C) and put in column (C) :
:-----:

```

If a number in the first column is not equal to (K), it is put in the second column. If a number in the first column is equal to (K), zero is put at the bottom of the second column. NRMAX is reset to the number of rows retained in the second column.

```

/-----\
/ OMIT rows with numbers between (K) and (K) in column (C) and put in column (C) \
\-----/

```

If a number in the first column is not between (K) and (K), inclusive, it is put in the second column. If a number in the first column is in the interval (K) to (K), zero is put at the bottom of the second column. NRMAX is reset to the number of rows retained in the second column.

```

/-----\
/ OMIT rows with (K) in col (C), corresp. rows of (C),..., (C) put in (C),..., (C) \
\-----/

```

The number of arguments in this form of OMIT is  $1 + 2n$ , where  $n = 1$ , or  $2$ , or  $3$ , ..., up to  $49$ . There are two sets of  $n$  column numbers. If (K) is found in any row of the first column of the first set of columns, that row of every column in the first set is not included in the second set of columns. Zeros are put at the bottom of the second set of columns and NRMAX is reset.

```

/-----\
/ OMIT from (K) to (K) in (C), corresp. rows of (C),..., (C) put in (C),..., (C) \
\-----/

```

This fourth form of OMIT bears the same relation to the second form as the the third form does to the first. The instruction has  $2 + 2n$  arguments, where  $n$  can be any number from  $1$  to  $48$ .

For the data in columns 1 to 4 shown below, the instructions

OMIT rows with -2.0 in column 1 and put in column 5  
 RESET 15  
 OMIT rows with numbers between -10.0 and 0.0 in column 1 and put in column 6  
 RESET 15  
 OMIT rows with -2.0 in col 1 corr row of col 2 put in cols 7 and 8  
 RESET 15  
 OMIT rows between -10.0 and 0.0 in col 1 corr rows of col 2 put in cols 9 and 10

would put data in columns 5 to 10 as shown in the following table:

Row/Column	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>8</u>	<u>9</u>	<u>10</u>
1	6	25	40	101	6	6	6	25	6	25
2	7	23	48	-1	7	7	7	23	7	23
3	7	25	40	104	7	7	7	25	7	25
4	6	25	49	109	6	6	6	25	6	25
5	-2	26	40	106	7	7	7	24	7	24
6	7	24	48	102	6	6	6	-3	6	-3
7	6	-3	40	103	6	6	6	25	6	25
8	6	25	49	108	6	6	6	23	6	23
9	6	23	-2	105	7	7	7	24	7	24
10	7	24	48	108	-1	6	-1	24	6	25
11	-1	24	48	102	6	7	6	25	7	26
12	6	25	49	101	7	7	7	26	7	25
13	7	26	48	-1	7	7	7	25	7	26
14	7	25	48	103	7		7	26		
15	7	26	49	107						

For the first OMIT instruction, there is a -2 in row 5 of column 1 and this row is omitted in column 5. NRMAX is reset to  $15 - 1 = 14$ . For the second OMIT instruction, the numbers in rows 5 and 11 in column 1 are between -10.0 and 0.0 and these rows are omitted in column 6. NRMAX is reset to  $15 - 2 = 13$ . For the third OMIT instruction, there is a -2 in row 5 of column 1 and hence row 5 is omitted from columns 7 and 8. NRMAX is reset to  $15 - 1 = 14$ . For the fourth (and last) OMIT instruction, the numbers in rows 5 and 11 in column 1 are between -10.0 and 0.0 and hence rows 5 and 11 of columns 1 and 2 are omitted in columns 9 and 10. NRMAX is reset to  $15 - 2 = 13$ .

```

: _____ :
: RECODE column (C) into column (C) :
: _____ :

```

The number 1 is assigned to the smallest number in the first column, the number 2 to the second smallest number, and so on and the result is stored in the column designated by the second argument. If the numbers 40, 49, 48, 40, 49, and 48 are in column 2, the instruction

RECODE column 2 into column 5

would put the numbers 1, 3, 2, 1, 3, and 2 into column 5.

```

: _____ :
: REPLACE the number (K) in column (C) by the number (K) and put in column (C) :
: _____ :

```

The replace instruction is another coding instruction which simply replaces one number by another. It can be used for the treatment of missing or outlying data. For the numbers -1, -9, 2, 5, and -9 in column 5, the instruction

REPLACE -9.0 in column 5 by 3.0 and put in column 7

would put the numbers -1, 3, 2, 5, and 3 into column 7.

REPLACE the numbers from (K) to (K) in col (C) by (K) and put in column (C)

This form of REPLACE is the same as the one above; except a range of numbers, instead of just one, is replaced by a single number. For the numbers -1, -9, 2, 5, and -9 in column 5, the instruction

REPLACE numbers from -3.0 to +3.0 in col 5 by 7.0 and put in col 9

would put the numbers 7, -9, 7, 5, and -9 into column 9.

RETAIN numbers between (K) and (K) in cols (C),..., (C) put in cols (C),..., (C)

Any row in the first set of columns with all numbers between (K) and (K), inclusive, is put at the top of the second set of columns. Any row in the first set of columns which has a number outside the range (K) to (K) is deleted. Zeros are put at the bottom of the second set of columns and NRMAX is reset to the number of rows retained. The instruction

RETAIN all numbers between 0.0 and 200.0 in cols 1,2,3, and 4 put in cols 11,12,13 and 14

would produce the following results for the numbers shown in columns 1 to 4:

Row/Column	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>11</u>	<u>12</u>	<u>13</u>	<u>14</u>
1	6	25	40	101	6	25	40	101
2	7	23	48	-1	7	25	40	104
3	7	25	40	104	6	25	49	109
4	6	25	49	109	7	24	48	102
5	-2	26	40	106	6	25	49	108
6	7	24	48	102	7	24	48	108
7	6	-3	40	103	6	25	49	101
8	6	25	49	108	7	25	48	103
9	6	23	-2	105	7	26	49	107
10	7	24	48	108				
11	-1	24	48	102				
12	6	25	49	101				
13	7	26	-1	-1				
14	7	25	48	103				
15	7	26	49	107				

In this example, the six rows 2, 5, 7, 9, 11, and 13 contain numbers in columns 1 to 4 which are outside the range 0.0 to 200.0 and are deleted. All the numbers in the remaining rows of columns 1 to 4 are within the range 0.0 to 200.0 and these rows are retained. NRMAX is reset to 6 = 15 - 9.

#### 4. STATISTICAL ANALYSIS.

##### 4.1\* Elementary Analysis.

If the specified starting point (K) of a FREQUENCY distribution instruction exceeded the smallest number in the data column, trouble developed. A correction has been made so that data below the starting point are ignored. If the numbers 1.0 to 10.0 are in column 2, the instruction:

FREQUENCY of col 2, using 3 classes of length 2.0, starting at 4.5, put in col 7  
will now store the frequency 2.0 in the first three rows of column 7 for the three classes:

4.5 to 6.5  
6.5 to 8.5  
8.5 to 10.5

Note, that the four measurements 1.0, 2.0, 3.0 and 4.0 are simply ignored and the total frequency count is 6 rather than 10. Furthermore, no diagnostic is given. One can deliberately set the starting point above the minimum to advantage, but care should be exercised in doing it.

Also, a similar situation exists if the specified number of classes (k) is not large enough to include all the data. For the above example, the instruction:

FREQUENCY of col 2, using 3 classes of length 2.0, starting at 0.5, put in col 8  
will now store the frequency 2.0 in the first three rows of column 8 for the three classes:

0.5 to 2.5  
2.5 to 4.5  
4.5 to 6.5

The total frequency count is six rather than ten, and the numbers 7.0, 8.0, 9.0 and 10.0 are ignored. Again, no diagnostic is given.

If the number of classes exceeds the number of rows in the worksheet, the following informative diagnostic is printed:

COLUMN NOT LONG ENOUGH TO STORE ALL NUMBERS.  
FIRST (n) NUMBERS WERE STORED.

The HISTOGRAM and NHISTOGRAM instructions have been revised to allow printing on narrow carriage terminals or CRT devices. In addition, a scale is now printed below the column titles and three columns have been added to show the relative cumulative frequencies, cumulative frequencies and relative frequencies. Page 61 is the equivalent of page 114 of NBS TN 552. Page 62 shows the result of using the instructions:

WIDTH 72

FREQUENCY of COMBINED MPG using 11 cells, length 2.0, start 11.5 store 11 12 13  
HISTOGRAM 11, 12, 13

with the data on page 110.

HISTOGRAM WITH MID-POINTS IN COLUMN 44 , FREQUENCIES IN COLUMN 43

MID-POINTS	RCF	CF	RF	FREQUENCY
				0-----10-----20
-.2500000	.205	8	.205	8 ++++++
.2500000	.487	19	.282	11 ++++++
.7500000	.846	33	.359	14 ++++++
1.2500000	.974	38	.128	5 +++++
1.7500000	1.000	39	.026	1 +

RCF = RELATIVE CUMULATIVE FREQUENCY  
 CF = CUMULATIVE FREQUENCY  
 RF = RELATIVE FREQUENCY

MID-POINTS	RCF	CF	RF	FREQUENCY
				0-----10-----20
-.2500000	.205	8	.205	8 ++++++
.2500000	.487	19	.282	11 ++++++
.7500000	.846	33	.359	14 ++++++
1.2500000	.974	38	.128	5 +++++
1.7500000	1.000	39	.026	1 +

RCF = RELATIVE CUMULATIVE FREQUENCY  
 CF = CUMULATIVE FREQUENCY  
 RF = RELATIVE FREQUENCY

LIST OF DATA, INSTRUCTIONS AND DIAGNOSTICS

SET VELOCITY OF LIGHT IN COLUMN 1

.4	.6	1.0	1.0	1.0	.5	.6	.7	1.0	.6	.2	1.9	.2
.4	.0	-.4	-.3	.0	-.4	-.3	.1	-.1	.2	-.5	.3	-.1
.2	-.2	.8	.5	.6	.8	.7	.7	.2	.5	.7	.8	1.1

\* INFORMATIVE DIAGNOSTIC FOR THE ABOVE INSTRUCTION -  
 NRMAX HAS BEEN RESET FROM 0 TO 39.

FREQUENCY OF COL 1 USING 5 CELLS OF LENGTH .5 START AT -.5 PUT IN COLS 41,42,43

\* INFORMATIVE DIAGNOSTIC FOR THE ABOVE INSTRUCTION -  
 NRMAX HAS BEEN RESET FROM 39 TO 5.

ADD COL 41 TO COL 42, MULT BY 0.5, ADD 0.0 AND PUT IN COL 44  
 HISTOGRAM WITH MID-POINTS IN COL 44 AND FREQUENCIES IN COL 43  
 SPACE 2  
 NHISTOGRAM WITH MID-POINTS IN COL 44 AND FREQUENCIES IN COL 43

HISTOGRAM WITH FREQUENCIES IN COLUMN 13

MID-POINTS	RCF	CF	RF	FREQUENCY
				0-----30-----60-----90
12.500000	.105	40	.105	40 ++++++
14.500000	.321	122	.216	82 ++++++
16.500000	.455	173	.134	51 ++++++
18.500000	.579	220	.124	47 ++++++
20.500000	.721	274	.142	54 ++++++
22.500000	.803	305	.082	31 ++++++
24.500000	.868	330	.066	25 ++++++
26.500000	.934	355	.066	25 ++++++
28.500000	.982	373	.047	18 ++++++
30.500000	.987	375	.005	2 +
32.500000	1.000	380	.013	5 ++

RCF = RELATIVE CUMULATIVE FREQUENCY

CF = CUMULATIVE FREQUENCY

RF = RELATIVE FREQUENCY

Previously, a plus sign (+) was printed for each measurement. If there was not enough room to print all the plus signs for a given class on one line, the printing continued on additional lines. Now, printing for any class of the distribution is on a single line and the scale is adjusted appropriately. Each plus sign may represent more than one measurement as indicated by the scale at the top. The zero of the number printed in the scale is printed at the 0th, 10th, 20th, etc. marks on the scale. In the second example discussed above, each tenth mark on the scale is 0, 30, 60 and 90. Hence, each plus represents three measurements; except the last plus may represent only one or two measurements.

Optional forms of HISTOGRAM and NHISTOGRAM now exist with three arguments instead of two, which specify the class boundaries of the frequency distribution instead of the mid-points. However, the mid-points are still printed rather than the class boundaries.

```
HISTOGRAM for class boundaries in columns (C) and (C) frequencies in column (C)
```

```
NHISTOGRAM for class boundaries in columns (C) and (C) frequencies in column (C)
```

#### 4.1.1 Stem-and-Leaf Displays.

STEM LEAF, SSTEM LEAF

A STEM LEAF instruction produces a stem-and-leaf graphical display of data; an innovation of Tukey (1970). The display is a concise representation of a set of measurements which is similar to a histogram, but has certain advantages. The stem-and-leaf display preserves more of the fine structure of the data and obviates the problem of choosing an appropriate class width for the histogram. Additional information is given on order statistics, such as the median, which makes it a better technique for studying the frequency distribution of measurement data in order to formulate mathematical models and to evaluate "outlying" values and other anomalies, so that summary measures (average, standard deviation) can be correctly interpreted.

Very simply, a stem-and-leaf display replaces the mark (X) of a histogram by a digit, thereby preserving the graphical features of the histogram and at the same time providing more information. The three measurements 103, 107 and 108 might appear in a histogram as

100-109 XXX

The same three measurements could appear in a stem-and-leaf display as

10 | 378

The stem, 10, gives the first two digits of the measurements and the leaves 3, 7 and 8 give the third digit. The technique has been generalized in a number of ways to provide displays for all kinds of data.

There is a main instruction which uses an automatic rule to produce three, or sometimes two, displays without any storage. Three, rather than one, displays are given partly because it is not possible to automatically determine the best display and partly to allow one to look at the data in different ways. Two options enable certain storage. The automatic rule is good, but not infallible and sometimes does not produce the best display. Also, there are times when one wants a specific display not produced by the automatic rule. Hence, there are three other options to allow one to override the automatic rule and store results. The four forms SSTEM LEAF suppress the automatic display.

The main instruction is simply

```

:-----:
:  STEM LEAF of column (C)  :
:-----:

```

Suppose column 7 contains the 114 measurements:

116.	82.	127.	100.	94.	117.	90.	121.	93.	120.	83.	95.	88.	123.
81.	75.	100.	98.	88.	86.	80.	105.	117.	92.	81.	85.	90.	100.
106.	98.	93.	96.	90.	108.	138.	133.	83.	100.	85.	95.	98.	123.
109.	108.	119.	100.	101.	129.	80.	92.	119.	94.	104.	91.	96.	103.
131.	115.	104.	103.	116.	103.	133.	77.	65.	80.	134.	103.	104.	86.
81.	89.	115.	130.	149.	126.	85.	93.	126.	93.	93.	82.	95.	89.
83.	124.	95.	84.	82.	75.	97.	100.	83.	83.	77.	78.	125.	113.
104.	121.	105.	80.	92.	103.	116.	68.	86.	76.	102.	94.	141.	86.
126.	119.												

The instruction

STEM LEAF of column 7

will cause the three displays shown on pages 65 and 66 to be printed. An explanation of a display follows.

Note, a careful examination of the stem-and-leaf display shows that some digits seem to occur more frequently than expected and some less frequently. In the third line of the first display, there are several 0's, 3's and 6's, but only one 4 and no 7's. Further investigation would be needed to determine whether the differences in expected frequencies are real or simply due to chance, but in any case a histogram would not reveal any difference.

#### Automatic Display

Each display starts on a new page which begins with a title

STEM AND LEAF DISPLAY FOR COLUMN XX

where XX is a column number heading or label. The display has four columns for the (i) depth, (ii) scrawl, (iii) stem and (iv) leaves. At the bottom of the display, additional information is printed: (v) numerical values for the scrawl, (vi) units of measurement for the stem and leaves with examples, and (vii) values of the four parameters used to determine the form of the display.

STEM AND LEAF DISPLAY FOR COLUMN 7

DEPTH	SCRAWL	STEM	LEAF
2	*	6	( 58
8		7	) 556778
35	H	8	( 000011122233333455566668899
	M	9	) 0001222333334445555667888
54		10	( 00000012333334444556889
31	H	11	) 35566677999
20		12	( 011334566679
8		13	) 013348
2	*	14	( 19

SCRAWL (N= 114, \*= 65.000000, H= 86.000000,  
M= 97.500000, H= 115.00000, \*= 149.00000)

EACH STEM UNIT IS 10  
EACH LEAF UNIT IS 1  
6 ( 5 READS AS 65  
14 ( 9 READS AS 149

FOR THE ABOVE STEM AND LEAF DISPLAY, THE FOLLOWING VARIABLES WERE USED  
I = 1, J = 2, K = 2, L = 1

STEM AND LEAF DISPLAY FOR COLUMN 7

DEPTH	SCRAWL	STEM	LEAF
2	*	6	( 58
2		7	)
8		7	( 556778
24		8	) 0000111222333334
35	H	8	( 55566668899
50		9	) 000122233333444
	M	9	( 5555667888
54		10	) 00000012333334444
37		10	( 556889
31		11	) 3
30	H	11	( 5566677999
20		12	) 011334
14		12	( 566679
8		13	) 01334
3		13	( 8
2		14	) 1
1	*	14	( 9

SCRAWL (N= 114, \*= 65.000000, H= 86.000000,  
M= 97.500000, H= 115.00000, \*= 149.00000)

EACH STEM UNIT IS 10  
EACH LEAF UNIT IS 1  
6 ( 5 READS AS 65  
14 ( 9 READS AS 149

FOR THE ABOVE STEM AND LEAF DISPLAY, THE FOLLOWING VARIABLES WERE USED  
I = 1, J = 2, K = 4, L = 1

## STEM AND LEAF DISPLAY FOR COLUMN 7

DEPTH	SCRAWL	STEM	LEAF
1	*	6	( 5
1		6	)
2		6	( 8
2		7	)
2		7	(
4		7	) 55
7		7	( 677
8		7	) 8
15		8	( 0000111
23		8	) 22233333
27		8	( 4555
31	H	8	) 6666
35		8	( 8899
39		9	) 0001
47		9	( 22233333
54		9	) 4445555
57		9	( 667
	M	9	) 888
54		10	( 0000001
47		10	) 233333
41		10	( 444455
35		10	) 6
34		10	( 889
31		11	)
31		11	( 3
30	H	11	) 55
28		11	( 66677
23		11	) 999
20		12	( 011
17		12	) 33
15		12	( 45
13		12	) 6667
9		12	( 9
8		13	) 01
6		13	( 33
4		13	) 4
3		13	(
3		13	) 8
2		14	( 1
1		14	)
1		14	(
1		14	)
1	*	14	( 9

SCRAWL (N= 114, \*= 65.000000, H= 86.000000,  
M= 97.500000, H= 115.00000, \*= 149.00000)

EACH STEM UNIT IS 10  
EACH LEAF UNIT IS 1  
6 ( 5 READS AS 65  
14 ( 9 READS AS 149

FOR THE ABOVE STEM AND LEAF DISPLAY, THE FOLLOWING VARIABLES WERE USED  
I = 1, J = 2, K = 10, L = 1

(i) Depth column: The depth is basically a cumulative frequency count, but the counting works both from the minimum down to the middle and from the maximum up to the middle of the display. From the minimum to the median (top half of the display), the depth for any stem is the total number of measurements less than or equal to the measurement corresponding to the largest leaf. On the second line of the first display in the above example, the stem is 7 and the largest leaf is 8. This represents the measurement 78. The depth, 8, is the total number of measurements less than or equal to 78. From the maximum to the median (bottom half of display), the depth for any stem is the total number of measurements equal to or greater than the measurement corresponding to the smallest leaf. On the seventh line of the first display in the above example, the stem is 12 and the smallest leaf is 0. Thus, there are 20 measurements equal to or greater than 120. A value for the depth is not given for the stem where the median lies. In some cases, it will take more than one line to print all the leaves for a particular stem. When this occurs, the depth is printed on the last line for that particular stem.

Nemenyi et al. (1975), give an approximate 95 percent confidence interval for the population median of a set of measurements which can be easily determined from the depth of a stem-and-leaf display. Let  $a$  be the integral part of  $(n+1)/2 - \sqrt{n}$ , where  $n$  is the number of measurements. In our example  $n = 114$  and  $a$  is the integral part of  $57.5 - 10.7 = 46.8$  or  $a = 46$ . Now find the measurements corresponding to the depth  $a = 46$  to obtain the 95 percent confidence interval. Thus, we can be 95 percent confident the median of our distribution is between 93 and 102.

(ii) Scrawl column: The five symbols used are:

- \* for the minimum, or smallest measurement
- H for the lower hinge, or lower quartile
- M for the median
- H for the upper hinge, or upper quartile
- \* for the maximum, or largest measurement

The lower hinge is that value such that 25 percent of the measurements fall below it and 75 percent of the measurements fall above it. Similarly, the upper hinge is that value such that 75 percent of the measurements fall below it and 25 percent of the measurements fall above it. The median, as usual, is the value for which 50 percent of the measurements fall below it and 50 percent of the measurements fall above it.

More precisely, let  $r = 1, 2$  and  $3$  for the lower hinge, median and upper hinge and  $x(r)$  be the measurements ordered from smallest to largest. Let  $m$  be the largest integer less than or equal to  $rn/4$ , where  $n$  is the number of measurements. If  $rn/4$  is not an integer use  $X(m+1)$  to define the lower hinge, median and upper hinge. If  $r*n/4$  is an integer, use  $[X(m) + x(m+1)]/2$ .

Each symbol is printed on the line in which the appropriate leaf appears. In the above example, the median is 97.5. For this measurement, the stem is 9 and the leaf 8. Thus, M is printed on the fourth line of the first display opposite the 23rd leaf, 7.

(iii) Stem column: The stem gives the leading significant digits of a measurement.

In the first display of the above example, each stem is different ( $k = 2$ ).

In the second display, two lines have the same stem ( $k = 4$ ). The first is used for leaves with values 0, 1, 2, 3 or 4 and the second for leaves with the values 5, 6, 7, 8 or 9. Thus, the six numbers in the 130's, namely, 130, 131, 133, 133, 134 and 138, which are represented on one line in the first display are represented on two lines in the second display.

In the third display, five lines have the same stem ( $k = 10$ ). The first is used for leaves with values 0 or 1, the second for leaves with values 2 or 3, the third for 4 or 5,

the fourth for 6 or 7 and the fifth for 8 or 9. Thus, five lines are used to represent the six numbers in the 130's.

If there is more than one stem on a line, the stems are separated by a slash (/) and the corresponding sets of leaves are separated by a slash. Thus

4/5/6

could be used to display all numbers from 40 to 69 and

4/5/6 ( //3

would represent 63.

(iv) Leaf column: The leaves are separated from their stems by alternating left and right parentheses. (A vertical line symbol, which would be preferable, is not available on some computers.) The leaf is the next significant digit after the last significant digit in the stem. If a leaf has more than one digit (see below), the leaves are separated by a comma. This is also true for mixed displays when any of the leading digits are actually part of the stem (see below).

(v) Scrawl: Values of the smallest measurement, lower hinge, median, upper hinge and largest measurement are printed after giving the total number of measurements. In the above example, the smallest of the 114 measurements is 65 and the largest is 149. The lower hinge is 86.0, the median is 97.5, and the upper hinge is 115.0.

(vi) Scale: Below the scrawl, the units of measurement for the stems and leaves are printed and examples are given for the smallest and largest measurements.

(vii) Parameters: The values of four parameters i, j, k and l uniquely determine the form of a display and are printed at the bottom of each display. The values can be set as in three of the options described below.

#### Options.

The following two instructions are identical to the main STEM LEAF instruction, except they allow one to store the scrawl (or the scrawl and the depth) in the worksheet.

STEM LEAF of column (C) and put scrawl in column (C)

The scrawl will be stored in the second named column. The eight values stored are.

<u>Row</u>	<u>Description</u>
1	NRMAX
2	minimum value of data
3	the next value after the minimum
4	the lower hinge
5	the median
6	the upper hinge
7	the value preceding the maximum
8	maximum value of the data

STEM LEAF of col (C), scrawl in (C) and put depth in column (C)

Both the depth and the scrawl are stored in the specified columns. The first row of the column, where the depth information is stored, contains the total number of stems. The depth of the first stem line is in row 2, depth of the second stem line in row 3 and so forth. If the number of stems is greater than the number of rows, only the number of rows minus one depths will be stored.

The next three instructions are similar to the three preceding ones, except the user defines the parameters i, j, k and  $\ell$ .

STEM LEAF of column (C) for parameter values (i), (j), (k) and ( $\ell$ )

STEM LEAF of col (C) for (i), (j), (k) and ( $\ell$ ) put scrawl in col (C)

STEM LEAF of col (C) for (i), (j), (k), ( $\ell$ ) put scrawl in (C) and depth in (C)

(i): The parameter i determines whether a simple display is given,  $i > 0$ , or a mixed display is given,  $i = 0$ .

Mixed stem-and-leaf displays, which are necessarily more complicated, can be used to provide a more concise display when the data vary over several orders of magnitude.

The numbers 78, 85, 96, 102 and 113 can be easily shown with a simple stem-and-leaf display as:

```
7 ( 8
8 ) 5
9 ( 6
10 ) 2
11 ( 3
```

The numbers 78, 85, 96, 198, 276 could not be displayed so easily without using a lot of stems (10, 11, ..., 18 and 20, 21, ..., 26) which had no leaves. But a mixed stem-and-leaf display would look like:

```
7 ( 8
8 ) 5
9 ( 6
1* ) 98
2* ( 76
```

The asterisk after the 1 (or 2) means that the first digit of the leaf is really part of the stem. Thus, 1\* ( 98 reads as 198, whereas 1 ( 98 reads as 19 and 18. If two asterisks are used, then the first two digits of the leaf are actually part of the stem, etc.

(j): The parameter j is the number of significant digits in the largest stem. It is increased by one, if the decimal point falls within the stem. The value of j must be

greater than zero. In the first display of the original example, the smallest stem is 6 and the largest is 14 with two digits and hence  $j = 2$ .

(k): The parameter  $k$  controls the length of each stem. If the display is simple ( $i > 0$ ), then  $k = 1, 2, 4$  or  $10$ . If  $k = 2$ , each stem is on a different line. If  $k = 4$ , a stem will appear twice. If  $k = 10$ , a stem will appear five times. Thus, the numbers 61, 68 and 73 (or 610, 680 and 730) would be displayed as:

for  $k = 2$ , and as

$$\begin{array}{l} 6 ( 18 \\ 7 ) 3 \end{array}$$

for  $k = 4$ , and as

$$\begin{array}{l} 6 ( 1 \\ 6 ) 8 \\ 7 ( 3 \end{array}$$

$$\begin{array}{l} 6 ( 1 \\ 6 ) \\ 6 ( \\ 6 ) \\ 6 ( 8 \\ 7 ) \\ 7 ( 3 \end{array}$$

for  $k = 10$ . In other words, the unit of measurement is 1, .5 and .2 times some power of ten depending upon whether  $k = 2, 4$  or  $10$ .

If  $k = 1$ , there will be two stems per line; each separated by a slash. The above three numbers, 61, 68 and 73 could be displayed as:

$$6/7 ( 18/3$$

Note, that the leaves for each stem are separated by a slash.

If the display is mixed,  $i = 0$ , then  $k = 1$  or  $3$ . If  $k = 1$ , there will be three stems per line as in:

$$6/7/8 ($$

If  $k = 3$ , each stem will be on a new line as in:

$$\begin{array}{l} 6 ( 18/3/ \\ 7 ) 3 \\ 8 ( \end{array}$$

( $\ell$ ): The value of the parameter  $\ell$  determines the number of digits in any leaf. For the instruction

STEM LEAF of col 7 for  $i=1, j=2, k=1$ , and  $\ell=1$

only one display form is printed. This display contains two stem values per stem line.

STEM AND LEAF DISPLAY FOR COLUMN 7

DEPTH	SCRAWL	STEM	LEAF
8	*	6/ 7 (	58/556778
	HM	8/ 9 )	0000111222333334455566668899/000122233333444455556667888
54	H	10/11 (	00000012333334444556889/35566677999
20		12/13 )	011334566679/013348
2	*	14/15 (	19/

SCRAWL (N= 114, \*= 65.000000, H= 86.000000,  
M= 97.500000, H= 115.000000, \*= 149.00000 )

EACH STEM UNIT IS 10  
EACH LEAF UNIT IS 1  
6/7 ( 5 READS AS 65  
14/15 ( 9 READS AS 149

FOR THE ABOVE STEM AND LEAF DISPLAY, THE FOLLOWING VARIABLES WERE USED  
I = 1, J = 2, K = 1, L = 1

Suppress Display.

The four forms of the SSTEM LEAF instruction are the same as the second, third, fifth and sixth forms of the STEM LEAF instruction except the automatic display is suppressed.

```
.....  
: SSTEM LEAF of column (C) put scrawl in column (C) :  
: : :  
: : :  
.....
```

```
.....  
/ SSTEM LEAF of col (C) put scrawl in col (C) and depth in col (C) :  
.....
```

```
.....  
/ SSTEM LEAF col (C) for (i), (j), (k) and (l), put scrawl in col (C) :  
.....
```

```
.....  
/ SSTEM LEAF of (C) for (i), (j), (k) and (l), put scrawl in (C) and depth in (C) :  
.....
```

Diagnostics:

Three fatal errors may occur in the use of the STEM LEAF or SSTEM LEAF instructions. If NRMAX is less than 5 or all the data values are the same, the following fatal error will be given:

NRMAX IS LESS THAN 5 OR ALL MEASUREMENTS ARE THE SAME.

Whenever the range of data is too large, the fatal error

VALUE OF SOME MEASUREMENT IS NOT ACCEPTABLE.

is printed. The most frequent fatal error occurs if any of the parameter values i, j, k and/or l are not defined correctly, or if the rules for determining the parameters are not satisfied. The error message given for these difficulties is:

I, J, K AND/OR L NOT DEFINED CORRECTLY, OR RULE  
CANNOT DETERMINE PROPER PARAMETER VALUES FOR THESE DATA.

The informative diagnostics which can occur with STEM LEAF and SSTEM LEAF instructions are:

PRINTING OF STEM AND LEAF DISPLAY IS NOT POSSIBLE.  
INSTRUCTION WAS TREATED AS SSTEM LEAF.

This happens if the stem (j) is very large and the defined printer line width is small.

(n) ROWS IN WORKSHEET IS NOT ENOUGH FOR COMPLETE STORAGE.

and

DISPLAY IS (n) LINES, ONLY FIRST 99 DISPLAYED.

If the parameters i, j, k and l are not defined carefully, the total number of stems may be huge.

#### Rule.

If the parameters are not defined by the user, STEM LEAF and SSTEM LEAF instructions will set i and l equal to one and determine the best pair of j, k by one of the following rules.

##### Primary Rule.

Let  $A =$  number of display lines from lower hinge to upper hinge divided by the total number of lines in display.

Let  $B =$  total number of lines in display divided by the number of measurements.

Then choose j and k such that  $.2 \leq A < .5$ ,  $B < .15$  and  $k > 2$ . Otherwise stop whenever (i)  $j + l + 1$  is greater than the number of significant digits in the data; (ii)  $B \leq .03$ ; or (iii) total number of display lines is greater than  $50(\log_{10} n)$ .

##### Secondary Rule.

If the primary rule is not satisfied, choose the first pair of j, k such that k is greater than one and the total number of display lines is greater than or equal to 30.

##### Final Rule.

Whenever the primary and secondary rules fail, the best j, k pair will be selected which satisfies the conditions  $k = 4$  and the number of lines in the display is maximum but less than 30.

#### 4.1.2 Statistical Plots.

##### STATPLOTS

```
-----:
: STATPLOTS of column (C) :
:                               :
:-----:

```

A STATPLOTS instruction produces four very useful plots for an exploratory analysis of data. The plots can be very helpful for detecting any anomalies in the data such as nonrandomness, lack of independence, presence of outliers, etc.

Plot 1, in the upper left-hand corner, is a simple plot of the measurements  $X(i)$  vs. the row number  $i$ , or the order the measurements are entered into the worksheet. This plot may be used to detect many different patterns of non-randomness such as trends, outliers, etc.

Plot 2, in the upper right-hand corner, is an autoregressive plot of  $X(i)$  versus  $X(i-1)$ . The plot is particularly useful for assessing lack of independence in the data. If the measurements are well behaved, the plot should not show any recognizable pattern.

Plot 3, in the lower left-hand corner, is a histogram. The histogram, designed by J. J. Filliben, is somewhat different from a conventional histogram. The interval for the horizontal axis is 0.2 standard deviations so that the horizontal axis goes from -5 to +5 standard deviations. Two scales are shown: one for the values of the measurements and one below it for multiples of standard deviations from the mean. Frequency is shown on the vertical axis. Information is printed below the histogram which gives the number of measurements, the value of the 0.2 standard deviation class width, the number of observations in excess of the mean plus or minus 5 standard deviations, and the number of measurements represented by a plotted  $X$  in the histogram.

Plot 4, in the lower right-hand corner, is a normal probability plot. Section C4.7.1 gives a description of the advantages and uses of a probability plot. An assumption of a normal distribution of measurements is not satisfied, if the points do not lie approximately on a straight line. Any marked curvature indicated by the points in the plot is a sign of non-normality. Values of the probability plot correlation coefficient (discussed in Section C4.7.1), the scrawl (discussed in Section C4.1.1), the mean and the standard deviation are printed at the bottom of the plot.

The vertical and horizontal scales for all four plots are determined by the minimum and maximum values plotted. The plotting symbol for plots 1, 2 and 4 is a period (.). If two or more points coincide, the plotting symbol becomes an asterisk (\*). The plotting symbol for plot 3 (histogram) is an  $X$ . Depending upon the frequency indicated on the vertical axis, any  $X$  may represent 1 or more measurements. The maximum number of measurements represented by an  $X$  is given in the last line of information at the bottom of the histogram.

#### Effect of WIDTH and LENGTH

A LENGTH instruction has no effect on a STATPLOTS instruction. If the width, as determined by a WIDTH, INTERACTIVE or TERMINAL instruction, is less than 120, then plots 1 and 2 are printed on one page and plots 3 and 4 are printed on the next page. The information printed below the normal probability plot appears below plot 2 on the first page. The information printed below the histogram is printed below the normal probability plot on the next page.

#### Remark

Serious consideration was given to making the results of STATPLOTS part of the automatic printing of a STATISTICAL analysis instruction. It was finally decided that there should be a separate instruction. However, the user of a STATISTICAL analysis would be well advised to use a STATPLOTS instruction in conjunction with a STATISTICAL analysis instruction in most instances.

#### Remark

Four plots on one page are far more effective than four plots on separate pages. By having all the plots on one page the eye can survey all plots simultaneously. A plot may display evidence not obvious in the other plots. Sometimes one plot can be used to support evidence found in another plot. Occasionally, the evidence in two plots may seem to conflict. This frequently suggests that further analysis or experimentation is needed.

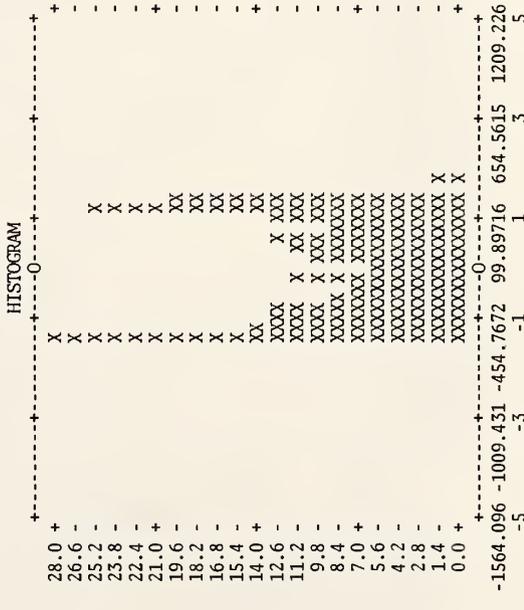
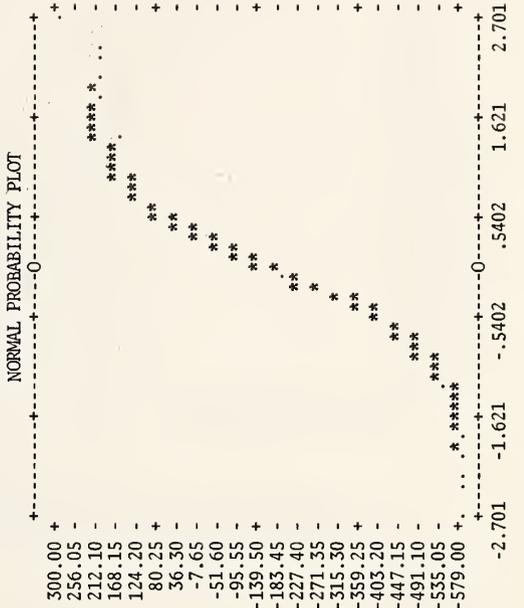
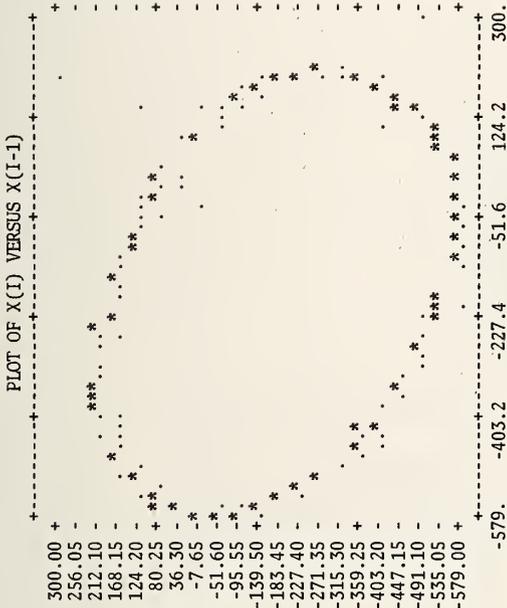
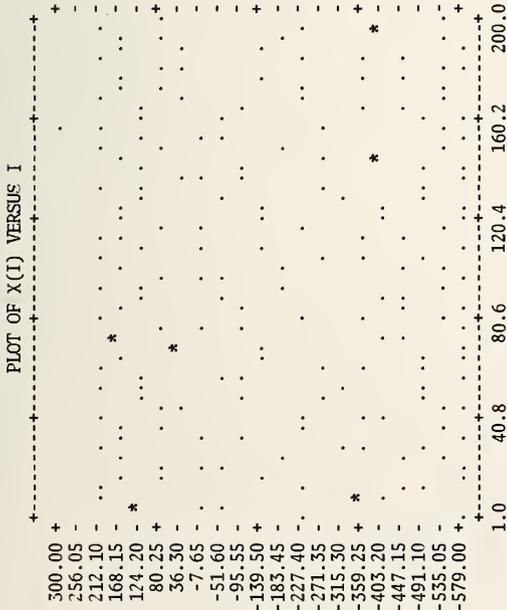
## Restrictions

1. The number of measurements (NRMAX) must be at least 5.
2. The number of measurements (NRMAX) may not exceed 6250.

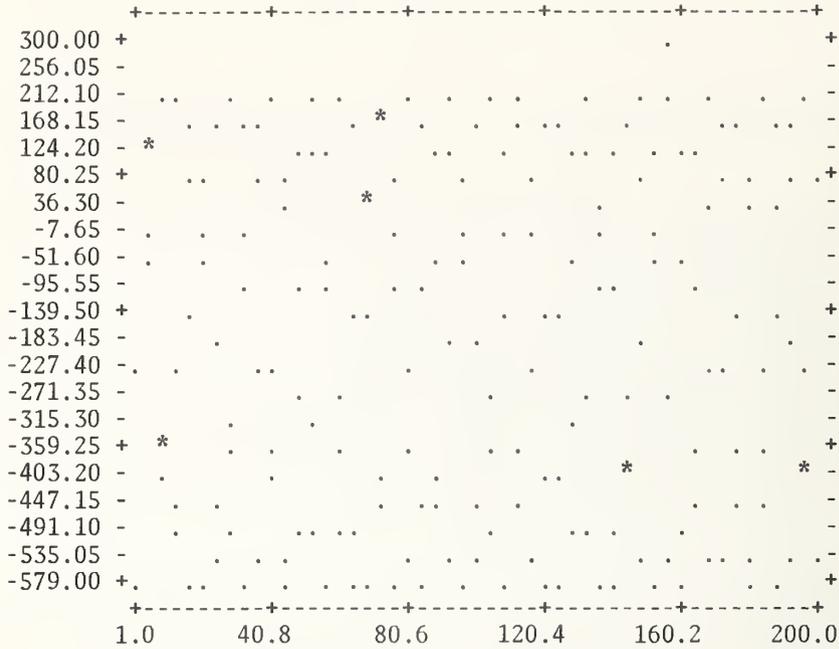
## Example

The set of instructions listed below is used to show the use of STATPLOTS to analyze a set of 200 steel beam deflection measurements. The results are show on pages 75, 76 and 77.

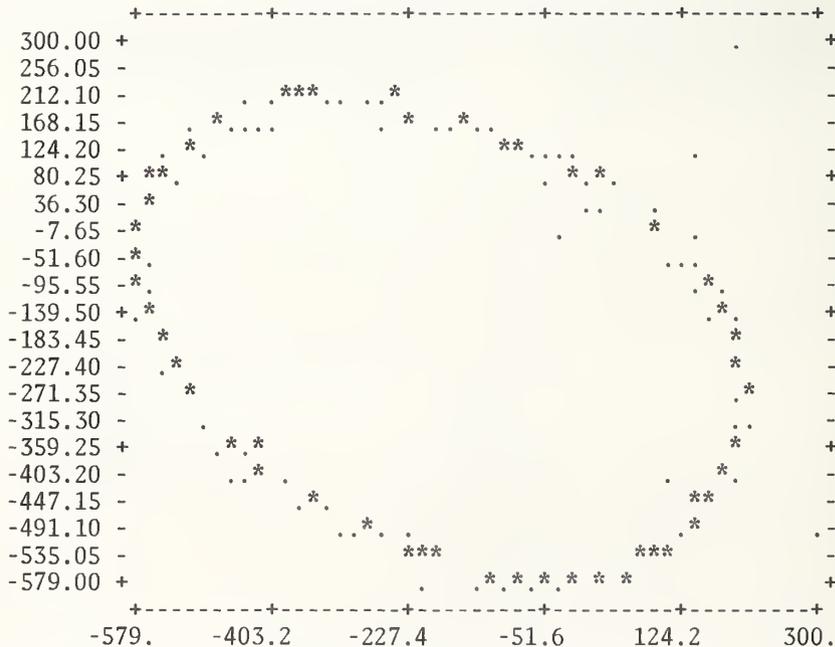
```
SET BEAM DEFLECTION DATA IN COLUMN 1
-213 -564 -035 -015 +141 +115 -420 -360 +203 -338
-431 +194 -220 -513 +154 -125 -559 +092 -021 -579
-052 +099 -543 -175 +162 -457 -346 +204 -300 -474
+164 -107 -572 -008 +083 -541 -224 +180 -420 -374
+201 -236 -531 +083 +027 -564 -112 +131 -507 -254
+199 -311 -495 +143 -046 -579 -090 +136 -472 -338
+202 -287 -477 +169 -124 -568 +017 +048 -568 -135
+162 -430 -422 +172 -074 -577 -013 +092 -534 -243
+194 -355 -465 +156 -081 -578 -064 +139 -449 -384
+193 -198 -538 +110 -044 -577 -006 +066 -552 -164
+161 -460 -344 +205 -281 -504 +134 -028 -576 -118
+156 -437 -381 +200 -220 -540 +083 +011 -568 -160
+172 -414 -408 +188 -125 -572 -032 +139 -492 -321
+205 -262 -504 +142 -083 -574 -000 +048 -571 -106
+137 -501 -266 +190 -391 -406 +194 -186 -553 +083
-013 -577 -049 +103 -515 -280 +201 +300 -506 +131
-045 -578 -080 +138 -462 -361 +201 -211 -554 +032
+074 -533 -235 +187 -372 -442 +182 -147 -566 +025
+068 -535 -244 +194 -351 -463 +174 -125 -570 +015
+072 -550 -190 +172 -424 -385 +198 -218 -536 +096
STATPLOTS OF COLUMN 1
WIDTH 80
STATPLOTS OF COLUMN 1
```



PLOT OF X(I) VERSUS I



PLOT OF X(I) VERSUS X(I-1)



PROBABILITY PLOT CORRELATION COEFFICIENT = .9540  
 SCRAWL (N= 200, \*= -579.00000, H= -453.00000,  
 M= -162.00000, H= 94.000000, \*= 300.00000)  
 MEAN = -177.43500  
 STANDARD DEVIATION = 277.33216



### 4.1.3 Table Making.

NTABLE AVERAGE,	NTABLE FREQUENCY,	NTABLE MEDIAN,	NTABLE MAXIMUM,
NTABLE MINIMUM,	NTABLE PERCENTAGE,	NTABLE PROPORTION,	NTABLE RANGE,
NTABLE STDDEV,	NTABLE SUM,	TABLE AVERAGE,	TABLE FREQUENCY,
TABLE MAXIMUM,	TABLE MEDIAN,	TABLE MINIMUM,	TABLE PERCENTAGE,
TABLE PROPORTION,	TABLE RANGE,	TABLE STDDEV,	TABLE SUM

A collection of data is often classified according to several variables, each of which can assume discrete values. For example, people can be classified by sex, marital status, education, etc. Sex, marital status, education, etc. will be called variables. The values that a variable may have will be called levels. The levels must have numerical values. The numbers 0 and 1 may be used to denote the two levels of sex. The interpretation of the data is aided considerably by cross tabulation or the construction of a table. The most common table is one which clearly shows the frequency (or count) for every possible combination in the classification. The table will show, for example, how many of the persons in a study are male, divorced and have a college education. Less frequently, but still important, there is a response variable, i.e., salary, and a table of a statistic of the response variable is needed. For example, a table showing the average salary for divorced males with a college education, single females with an advanced degree, etc., may be needed.

A large and powerful set of instructions for constructing and printing tables (cross tabulation) has been added to OMNITAB, which should find wide use in many different areas. Each command consists of two words. The first word is either TABLE or NTABLE. The N before TABLE in NTABLE has the usual significance; i.e., the printing does not start on a new page and titles are not printed. The second word in the command may be any of the ten words AVERAGE, FREQUENCY, MEDIAN, MAXIMUM, MINIMUM, PERCENTAGE, PROPORTION, RANGE, STDDEV, or SUM. There is an optional form of each instruction which permits automatic storage.

#### Data for examples

Examples are given below which use the following instructions and data:

```
READ DATA INTO COLUMNS 7 5 27 50
$
6 25 40 101
7 23 48 107
7 25 40 104
6 25 49 109
6 26 40 106
7 24 48 102
6 26 40 103
6 25 49 108
6 23 49 105
7 24 48 108
7 24 48 102
6 25 49 101
7 26 48 108
7 25 48 103
7 26 49 107
```

```
WIDTH 80
HEAD COL 5/ AGE
TABLE FREQUENCY 3 WAY FOR DATA IN COLUMNS 7 5 AND 27
ADD 0.98765 TO COLUMN 27 AND PUT IN COLUMN 28
```

## Instructions

```
NTABLE AVERAGE (n) way with levels in cols (C),(C),..., (C) for data in (C)
```

```
NTABLE FREQUENCY (n) way with levels in cols (C),(C),..., (C)
```

```
NTABLE MAXIMUM (n) way, levels in cols (C),(C),..., (C) for data in col (C)
```

The result of using the instructions

```
SPACE 5
NTABLE MAXIMUM 3 WAY FOR DATA IN COLUMNS 7, 5, 28 and 50
```

is shown at the bottom of page 89.

```
NTABLE MEDIAN (n) way with levels in cols (C),(C),..., (C) for data in col (C)
```

The result of using the instructions

```
SPACE 5
NTABLE MEDIAN 3 WAY FOR DATA IN COLUMNS 7, 5, 28 and 50
```

is shown at the bottom of page 90.

```
NTABLE MINIMUM (n) way with levels in cols (C),(C),..., (C) for data in col (C)
```

```
NTABLE PERCENTAGE (n) way with levels in cols (C),(C),..., (C)
```

```
NTABLE PROPORTION (n) way with levels in cols (C),(C),..., (C)
```

The result of using the instructions

```
SPACE 5
NTABLE PROPORTION 3 WAY FOR DATA IN COLUMNS 7, 5, 28
```

is shown at the bottom of page 91.

```
NTABLE RANGE (n) way with levels in cols (C),(C),..., (C) for data in col (C)
```

```
:  
: NTABLE STDDEV (n) way with levels in cols (C),(C),...,(C) for data in col (C)  
:
```

The result of using the instructions

```
SPACE 5  
NTABLE STDDEV 3 WAY FOR DATA IN COLUMNS 7, 5, 28 and 50
```

is shown at the bottom of page 88.

```
:  
: NTABLE SUM (n) way with levels in cols (C),(C),...,(C) for data in col (C)  
:
```

The result of using the instructions

```
SPACE 5  
NTABLE SUM 3 WAY FOR DATA IN COLUMNS 7, 5, 27 and 50
```

is shown at the bottom of page 87.

```
:  
: TABLE AVERAGE (n) way with levels in cols (C),(C),...,(C) for data in col (C)  
:
```

The result of using the instruction

```
TABLE AVERAGE 3 WAY FOR DATA IN COLUMNS 7, 5, 27 and 50
```

is shown at the top of page 88.

```
:  
: TABLE FREQUENCY (n) way with levels in cols (C),(C),...,(C)  
:
```

The result of using the instruction

```
TABLE FREQUENCY 3 WAY FOR DATA IN COLUMNS 7, 5, and 27
```

is shown on page 82. To show the effect of varying the number of variables, the result of using the instruction

```
TABLE FREQUENCY 4 WAY FOR DATA IN COLUMNS 7, 5, 27 and 50
```

is shown on pages 83 and 84 and the results of using the instructions

```
TABLE FREQUENCY 2 WAY FOR DATA IN COLUMNS 5 and 27  
SPACE 3  
NTABLE FREQUENCY 1 WAY FOR DATA IN COLUMN 5
```

are shown on page 87. The result of using the instructions

```
WIDTH 50  
TABLE FREQUENCY 3 WAY FOR DATA IN COLUMN 7, 50 and 27
```

is shown on pages 85 and 86. In this case, the number of levels of the second variable, 9, is too large to print the table on a page with 50 characters per line. The first five

columns are printed on page 85 and the remaining four columns and the totals are printed on page 86.

```
TABLE MAXIMUM (n) way with levels in cols (C),(C),..., (C) for data in col (C)
```

```
TABLE MEDIAN (n) way with levels in cols (C),(C),..., (C) for data in col (C)
```

```
TABLE MINIMUM (n) way with levels in cols (C),(C),..., (C) for data in col (C)
```

The result of using the instruction

```
TABLE MINIMUM 3 WAY FOR DATA IN COLUMNS 7, 5, 27 and 50
```

is shown at the top of page 89.

```
TABLE PERCENTAGE (n) way with levels in cols (C),(C),..., (C)
```

The result of using the instruction

```
TABLE PERCENTAGE 3 WAY FOR DATA IN COLUMNS 7, 5, and 27
```

is shown at the top of page 91.

```
TABLE PROPORTION (n) way with levels in cols (C),(C),..., (C)
```

```
TABLE RANGE (n) way with levels in cols (C),(C),..., (C) for data in col (C)
```

The result of using the instruction

```
TABLE RANGE 3 WAY FOR DATA IN COLUMNS 7, 5, 27 and 50
```

is shown at the top of page 90.

```
TABLE STDDEV (n) way with levels in cols (C),(C),..., (C) for data in col (C)
```

```
TABLE SUM (n) way with levels in cols (C),(C),..., (C) for data in col (C)
```

## 3-WAY TABLE OF FREQUENCIES OF 15 MEASUREMENTS

(1) = COLUMN 7 , 2 LEVELS  
 (2) = AGE , 4 LEVELS  
 (3) = COLUMN 27 , 3 LEVELS

	(2)	23	24	25	26	TOTALS
(3)	(1)					
40	6	0	0	1	2	3
	7	0	0	1	0	1
48	6	0	0	0	0	0
	7	1	3	1	1	6
49	6	1	0	3	0	4
	7	0	0	0	1	1
TOTALS		2	3	6	4	15

4-WAY TABLE OF FREQUENCIES OF 15 MEASUREMENTS

- (1) = COLUMN 7 , 2 LEVELS
- (2) = AGE , 4 LEVELS
- (3) = COLUMN 27 , 3 LEVELS
- (4) = COLUMN 50 , 9 LEVELS

(4)	(3)	(2) (1)	23	24	25	26	TOTALS
101	40	6	0	0	1	0	1
		7	0	0	0	0	0
	48	6	0	0	0	0	0
7		0	0	0	0	0	
49	6	0	0	1	0	1	
	7	0	0	0	0	0	
102	40	6	0	0	0	0	0
		7	0	0	0	0	0
	48	6	0	0	0	0	0
7		0	2	0	0	2	
49	6	0	0	0	0	0	
	7	0	0	0	0	0	
103	40	6	0	0	0	1	1
		7	0	0	0	0	0
	48	6	0	0	0	0	0
7		0	0	1	0	1	
49	6	0	0	0	0	0	
	7	0	0	0	0	0	
104	40	6	0	0	0	0	0
		7	0	0	1	0	1
	48	6	0	0	0	0	0
7		0	0	0	0	0	
49	6	0	0	0	0	0	
	7	0	0	0	0	0	
105	40	6	0	0	0	0	0
		7	0	0	0	0	0
	48	6	0	0	0	0	0
7		0	0	0	0	0	
49	6	1	0	0	0	1	
	7	0	0	0	0	0	

CONTINUATION

(4)	(3)	(2) (1)	23	24	25	26	TOTALS
106	40	6	0	0	0	1	1
		7	0	0	0	0	0
	48	6	0	0	0	0	0
		7	0	0	0	0	0
	49	6	0	0	0	0	0
		7	0	0	0	0	0
107	40	6	0	0	0	0	0
		7	0	0	0	0	0
	48	6	0	0	0	0	0
		7	1	0	0	0	1
	49	6	0	0	0	0	0
		7	0	0	0	1	1
108	40	6	0	0	0	0	0
		7	0	0	0	0	0
	48	6	0	0	0	0	0
		7	0	1	0	1	2
	49	6	0	0	1	0	1
		7	0	0	0	0	0
109	40	6	0	0	0	0	0
		7	0	0	0	0	0
	48	6	0	0	0	0	0
		7	0	0	0	0	0
	49	6	0	0	1	0	1
		7	0	0	0	0	0
	TOTALS		2	3	6	4	15

3-WAY TABLE OF FREQUENCIES OF 15 MEASUREMENTS

(1) = COLUMN 7 , 2 LEVELS  
 (2) = COLUMN 50 , 9 LEVELS  
 (3) = COLUMN 27 , 3 LEVELS

		(2)	101	102	103	104	105
(3)	(1)						
40	6		1	0	1	0	0
	7		0	0	0	1	0
48	6		0	0	0	0	0
	7		0	2	1	0	0
49	6		1	0	0	0	1
	7		0	0	0	0	0
TOTALS			2	2	2	1	1

3-WAY TABLE OF FREQUENCIES OF 15 MEASUREMENTS  
CONTINUATION

(1) = COLUMN 7 , 2 LEVELS  
 (2) = COLUMN 50 , 9 LEVELS  
 (3) = COLUMN 27 , 3 LEVELS

(3)	(2)	106	107	108	109	TOTALS
(1)	(1)					
40	6	1	0	0	0	3
	7	0	0	0	0	1
48	6	0	0	0	0	0
	7	0	1	2	0	6
49	6	0	0	1	1	4
	7	0	1	0	0	1
TOTALS		1	2	3	1	15

2-WAY TABLE OF FREQUENCIES OF 15 MEASUREMENTS

(1) = AGE , 4 LEVELS  
 (2) = COLUMN 27 , 3 LEVELS

(2) (1)	40	48	49	TOTALS
23	0	1	1	2
24	0	3	0	3
25	2	1	3	6
26	2	1	1	4
TOTALS	4	6	5	15

1-WAY TABLE OF FREQUENCIES OF 15 MEASUREMENTS

(1) = AGE , 4 LEVELS

(1)	
23	2
24	3
25	6
26	4
TOTAL	15

3-WAY TABLE OF SUMS OF 15 MEASUREMENTS IN COLUMN 50

(1) = COLUMN 7 , 2 LEVELS  
 (2) = AGE , 4 LEVELS  
 (3) = COLUMN 27 , 3 LEVELS

(3) (2) (1)	23	24	25	26	TOTALS
40	6	*	*	101	209
	7	*	*	104	*
48	6	*	*	*	*
	7	107	312	103	108
49	6	105	*	318	*
	7	*	*	*	107
TOTALS	212	312	626	424	1574

\* 14 OF THE 24 CELLS IN THE TABLE ARE EMPTY.

3-WAY TABLE OF AVERAGES OF 15 MEASUREMENTS IN COLUMN 50

(1) = COLUMN 7 , 2 LEVELS  
 (2) = AGE , 4 LEVELS  
 (3) = COLUMN 27 , 3 LEVELS

		(2)	23	24	25	26
(3)	(1)					
40	6	*	*	101.0	104.5	
	7	*	*	104.0	*	
48	6	*	*	*	*	
	7	107.0	104.0	103.0	108.0	
49	6	105.0	*	106.0	*	
	7	*	*	*	107.0	

\* 14 OF THE 24 CELLS IN THE TABLE ARE EMPTY.

3-WAY TABLE OF STANDARD DEVIATIONS OF 15 MEASUREMENTS IN COLUMN 50

(1) = COLUMN 7 , 2 LEVELS  
 (2) = AGE , 4 LEVELS  
 (3) = COLUMN 28 , 3 LEVELS

		(2)	23	24	25	26
(3)	(1)					
40.988	6	*	*	.0000	2.1213	
	7	*	*	.0000	*	
48.988	6	*	*	*	*	
	7	.0000	3.4641	.0000	.0000	
49.988	6	.0000	*	4.3589	*	
	7	*	*	*	.0000	

\* 14 OF THE 24 CELLS IN THE TABLE ARE EMPTY.

3-WAY TABLE OF MINIMUMS OF 15 MEASUREMENTS IN COLUMN 50

(1) = COLUMN 7 , 2 LEVELS  
 (2) = AGE , 4 LEVELS  
 (3) = COLUMN 27 , 3 LEVELS

	(2)	23	24	25	26	TOTALS
(3)	(1)					
40	6	*	*	101	103	101
	7	*	*	104	*	104
48	6	*	*	*	*	*
	7	107	102	103	108	102
49	6	105	*	101	*	101
	7	*	*	*	107	107
TOTALS		105	102	101	103	101

\* 14 OF THE 24 CELLS IN THE TABLE ARE EMPTY.

3-WAY TABLE OF MAXIMUMS OF 15 MEASUREMENTS IN COLUMN 50

(1) = COLUMN 7 , 2 LEVELS  
 (2) = AGE , 4 LEVELS  
 (3) = COLUMN 28 , 3 LEVELS

	(2)	23	24	25	26	TOTALS
(3)	(1)					
40.988	6	*	*	101	106	106
	7	*	*	104	*	104
48.988	6	*	*	*	*	*
	7	107	108	103	108	108
49.988	6	105	*	109	*	109
	7	*	*	*	107	107
TOTALS		107	108	109	108	109

\* 14 OF THE 24 CELLS IN THE TABLE ARE EMPTY.

3-WAY TABLE OF RANGES OF 15 MEASUREMENTS IN COLUMN 50

(1) = COLUMN 7 , 2 LEVELS  
 (2) = AGE , 4 LEVELS  
 (3) = COLUMN 27 , 3 LEVELS

	(2)	23	24	25	26	TOTALS
(3)	(1)					
40	6	*	*	0	3	5
	7	*	*	0	*	0
48	6	*	*	*	*	*
	7	0	6	0	0	6
49	6	0	*	8	*	8
	7	*	*	*	0	0
TOTALS		2	6	8	5	8

\* 14 OF THE 24 CELLS IN THE TABLE ARE EMPTY.

3-WAY TABLE OF MEDIANS OF 15 MEASUREMENTS IN COLUMN 50

(1) = COLUMN 7 , 2 LEVELS  
 (2) = AGE , 4 LEVELS  
 (3) = COLUMN 28 , 3 LEVELS

	(2)	23	24	25	26
(3)	(1)				
40.988	6	*	*	101.0	104.5
	7	*	*	104.0	*
48.988	6	*	*	*	*
	7	107.0	102.0	103.0	108.0
49.988	6	105.0	*	108.0	*
	7	*	*	*	107.0

\* 14 OF THE 24 CELLS IN THE TABLE ARE EMPTY.

3-WAY TABLE OF PERCENTAGES OF 15 MEASUREMENTS

(1) = COLUMN 7 , 2 LEVELS  
 (2) = AGE , 4 LEVELS  
 (3) = COLUMN 27 , 3 LEVELS

	(2)	23	24	25	26	TOTALS
(3)	(1)					
40	6	.00	.00	6.67	13.33	20.00
	7	.00	.00	6.67	.00	6.67
48	6	.00	.00	.00	.00	.00
	7	6.67	20.00	6.67	6.67	40.00
49	6	6.67	.00	20.00	.00	26.67
	7	.00	.00	.00	6.67	6.67
TOTALS		13.33	20.00	40.00	26.67	100.00

3-WAY TABLE OF PROPORTIONS OF 15 MEASUREMENTS

(1) = COLUMN 7 , 2 LEVELS  
 (2) = AGE , 4 LEVELS  
 (3) = COLUMN 28 , 3 LEVELS

	(2)	23	24	25	26	TOTALS
(3)	(1)					
40.988	6	.0000	.0000	.0667	.1333	.2000
	7	.0000	.0000	.0667	.0000	.0667
48.988	6	.0000	.0000	.0000	.0000	.0000
	7	.0667	.2000	.0667	.0667	.4000
49.988	6	.0667	.0000	.2000	.0000	.2667
	7	.0000	.0000	.0000	.0667	.0667
TOTALS		.1333	.2000	.4000	.2667	1.0000

## Title

Each table has a one line title at the top which gives (i) the number of variables in the classification, (ii) the statistic tabulated, and (iii) the number of measurements (NRMAX). The coding scheme for the variables is shown below the title and includes the column number where the levels of the variable are stored and the number of levels for each of the variables. If a LABEL or HEAD instruction has been used, the heading is used instead of the column number.

## Use of asterisks

A distinction is made between zero and nothing for all the tables; except those for frequencies, percentages and proportions. When no data exists for a specific combination of levels of variables, an asterisk is printed in the table. In the table of standard deviations on page 88 a zero is printed when there is only one measurement (or all the measurements are the same), but an asterisk is printed when the frequency is zero. A note is printed at the bottom of the table showing the number of cells in the table which are empty.

## Row and column totals printed

Row and column totals are printed for the TABLE (or NTABLE) FREQUENCY, MINIMUM, MAXIMUM, PERCENTAGES, PROPORTIONS, RANGE, and SUM instructions. The totals are not printed for the AVERAGE, MEDIAN or STDDEV instructions. The row totals are not printed for 1-way tables and no totals are printed for 1-way tables when the variable has a single level. For 2-way tables the row totals can be used directly. For 3-way and higher tables, the row totals can be summed by hand to obtain the different marginal totals.

The meaning of a frequency total is obvious. The meaning of any of the totals in a table of ranges may not be quite so obvious. In the table of ranges on page 90, the number 5 in the first row of the row totals is the difference between the largest number in column 50 corresponding to level 40 in column 27, 106, and the smallest, 101.

## Number of digits printed

Two decimal places are used to print the entries in the table produced by a TABLE (or NTABLE) PERCENTAGE instruction and four decimal places are used for a TABLE (or NTABLE) PROPORTIONS instruction. For all the other instructions and the levels of the variables, the maximum number of digits printed is five. The number of digits needed to print a number is computed. If a number can be printed with less than five digits, it is printed with the minimum number of digits required. Note, the difference in the printing of the levels of column 27 (40, 48 and 49) and the levels of column 28 (40.988, 48.988 and 49.988).

## Space for table

The tables are printed compactly without any waste. The actual size will vary depending upon the number of digits needed to print the numbers. Since three decimal places are needed to print the levels in column 28 and none are needed to print the levels in column 27, more space is used to print the table on page 94 than the table on page 82.

## Number of arguments

The first argument in any of the instructions is always, (n), the number of variables in the classification. The remaining arguments are always column numbers. The instructions for tabulating frequencies, percentages and proportions have (n+1) arguments, or (n+2) arguments if storage is requested. The instructions for tabulating averages, maximums, medians, minimums, ranges, standard deviations and sums always have (n+2) arguments, or (n+3) arguments if storage is requested.

## Restriction

The value of the first argument, (n), must be greater than zero. It must be less than or equal to eight in the instructions which table frequencies, percentages and proportions. It must be less than or equal to seven in the instructions which table averages, maximums, medians, minimums, ranges, standard deviations and sums.

## Remarks

Often data has to be preprocessed before a table is constructed. There may be missing observations. A variable may need to be coded. OMNITAB has an extremely powerful set of instructions for manipulating data. The editing instructions described in section C3.5.1 can be particularly useful for preprocessing data for cross tabulation.

If the second variable has a large number of levels, the page may not be wide enough to print all the columns on one page as can be seen on pages 85 and 86. It may be desirable to make the second variable the one with the fewest (or a small) number of levels.

## Storage

Each of the ten TABLE and the ten NTABLE instructions has an optional form with one additional argument to specify automatic storage. Data is stored in (n+1+r) columns, starting with the last column number (last argument), where r is the number of levels of the second variable.

Page 94 shows the results of using the instructions:

```
ADD 0.98765 TO COLUMN 27 AND PUT IN COLUMN 28
TABLE FREQUENCY 3 WAY FOR DATA IN COLUMNS 7, 5 and 28 STORE 31
SPACE 3
NPRINT COLUMNS 31 *** 35
NPRINT COLUMNS 36 *** 38
```

Items stored in the first column are:

<u>Row</u>	<u>Description</u>
1	length of table
2	number of columns in table
3	total size of table
4	number of columns of stored data
5	starting column for storing levels
6	last column of stored data

The table is stored in the next (r) columns. The last (n) columns contain the number of levels in the first row and the values of the levels for each of the (n) variables in rows below the first row.

The first row of columns 36 to 38 contain the numbers 2, 4 and 3 which are the numbers of levels of the variables stored in columns 7, 5 and 28. The number 6 in row 1 of column 31 equals  $2 \times 3$  equals the number of rows in the table. The number 24 in row 3 of column 31 equals the product of the numbers of levels ( $2 \times 4 \times 3$ ), which is the size of the table. The number of columns of storage in row 4 of column 31 is  $8 = 1 + 4 + 3$ . Rows 5 and 6 of column 31 contain the starting column, 36, and the last column, 38, for storing the values of the levels of each of the variables.

3-WAY TABLE OF FREQUENCIES OF 15 MEASUREMENTS

(1) = COLUMN 7 , 2 LEVELS  
 (2) = AGE , 4 LEVELS  
 (3) = COLUMN 28 , 3 LEVELS

	(2)	23	24	25	26	TOTALS
(3)	(1)					
40.988	6	0	0	1	2	3
	7	0	0	1	0	1
48.988	6	0	0	0	0	0
	7	1	3	1	1	6
49.988	6	1	0	3	0	4
	7	0	0	0	1	1
TOTALS		2	3	6	4	15

6.000000	0.	0.	1.000000	2.000000
4.000000	0.	0.	1.000000	0.
24.000000	0.	0.	0.	0.
8.000000	1.000000	3.000000	1.000000	1.000000
36.000000	1.000000		3.000000	0.
38.000000				1.000000

2.000000	4.000000	3.000000
6.000000	23.000000	40.987650
7.000000	24.000000	48.987650
	25.000000	49.987650
	26.000000	

The structure of the optional forms which specify automatic storage is shown below.

NTABLE AVERAGE (n) way, levels in (C)...(C), data in (C) start storing in (C)

NTABLE FREQUENCY (n) way, levels in cols (C),(C),...,(C) start storing in (C)

NTABLE MAXIMUM (n) way, levels in (C)...(C) data in (C) start storing in (C)

NTABLE MEDIAN (n) way, levels in (C),...,(C) data in (C) start storing in (C)

NTABLE MINIMUM (n) way, levels in (C)...(C) data in (C) start storing in (C)

NTABLE PERCENTAGE (n) way, levels in cols (C),(C),...,(C) start storing in (C)

NTABLE PROPORTION (n) way, levels in cols (C),(C),...,(C) start storing in (C)

NTABLE RANGE (n) way, levels in (C),...,(C) data in (C) start storing in (C)

NTABLE STDDEV (n) way, levels in cols (C)...(C) data in (C) start storing in (C)

NTABLE SUM (n) way, levels in cols (C),...,(C) data in (C) start storing in (C)

TABLE AVERAGE (n) way, levels in (C),...,(C) data in (C) start storing in (C)

TABLE FREQUENCY (n) way, levels in cols (C),(C),...,(C) start storing in (C)

TABLE MAXIMUM (n) way, levels in (C),..., (C) data in (C) start storing in (C)

TABLE MEDIAN (n) way, levels in (C),..., (C) data in (C) start storing in (C)

TABLE MINIMUM (n) way, levels in (C),..., (C) data in (C) start storing in (C)

TABLE PERCENTAGE (n) way, levels in cols (C), (C),..., (C) start storing in (C)

TABLE PROPORTION (n) way, levels in cols (C), (C),..., (C) start storing in (C)

TABLE RANGE (n) way, levels in cols (C),..., (C) data in (C) start storing in (C)

TABLE STDDEV (n) way, levels in (C),..., (C) data in (C) start storing in (C)

TABLE SUM (n) way, levels in cols (C),..., (C) data in (C) start storing in (C)

#### 4.2\* Analysis Of One Column Of Data.

The automatic printing and storage of a STATISTICAL analysis instruction has been greatly improved. However, none of the nine forms of the instruction has been changed.

#### Printing.

The automatic printing has been greatly improved by (1) using readable printing rather than scientific notation, (2) limiting the printing to 72 characters per line so that the instruction can be used with a narrow carriage terminal, (3) deleting statistics of little or no value, (4) adding statistics of substantial value, and (5) changing the style.

Without the "OMNITAB" title line at the top of the page, pages 99 and 100 show the automatic printing for the example on page 115 of TN 552 ("OMNITAB II User's Reference Manual," National Bureau of Standards Technical Note 552). In effect, they replace pages 117 and 118 of TN 552. Below are the two lines which would replace the first line at the top after STATISTICAL ANALYSIS OF ..., if weights of 1.0 in column 7 were specified.

WEIGHTS IN COLUMN 7 OF LENGTH 39 WITH 39 NON-ZERO WEIGHTS  
ALL COMPUTATIONS ARE BASED ON OBSERVATIONS WITH NON-ZERO WEIGHTS

Page 101 shows the second page of printing for the same data when weights of 1.0 are specified.

Page 102 shows the first page of the printing for the example discussed in "A User's Guide to the OMNITAB Command 'STATISTICAL ANALYSIS'," NBS TN 756.

Note that HEAD COL 2/ PIKES PEAK was used and PIKES PEAK is used after "STATISTICAL ANALYSIS OF" instead of COLUMN 2.

### Changes in Automatic Printing

- (1) Statistics deleted (all from OTHER STATISTICS) are:

WTD SUM OF VALUES  
WTD SUM OF SQUARES  
WTD SUM OF DEVS SQUARED  
STUDENT'S T  
WTD SUM ABSOLUTE VALUES  
WTD AVE ABSOLUTE VALUES

- (2) Statistics added are:

FREQUENCY DISTRIBUTION OF LEAST SIGNIFICANT DIGIT (0,1,...,9)  
INTER-QUARTILE RANGE  
QUADRATIC COEFF., SIGNIFICANCE LEVEL  
AUTOCORRELATION COEFFICIENT  
95% DISTRIBUTION-FREE CONFIDENCE INTERVAL FOR THE MEDIAN  
TOLERANCE INTERVALS  
SECOND MINIMUM  
SECOND MAXIMUM  
(MEAN-MINIMUM)/STANDARD DEVIATION  
(MAXIMUM-MEAN)/STANDARD DEVIATION  
LOWER QUARTILE  
UPPER QUARTILE

- (3) Modifications include:

- (i) Column heading determined by HEAD COL is used in the first line.
- (ii) The number of distinct measurements is printed, if weights are not specified.
- (iii) The class interval for the frequency distribution is printed.
- (iv) Either the weighted mean or the mean is printed, depending upon whether weights are specified or not, but not both. Similarly, only one mid-mean, or trimmed mean, is printed.
- (v) The absolute value of the coefficient of variation is printed.
- (vi) For trend statistics and other tests for non-randomness, a statistic,  $T$ , expected value of the statistic,  $\mu(T)$ , and the standard deviation of the statistic,  $\sigma(T)$ , have been replaced by the statistic,  $T$ , and its  $z$  value, where

$$z = \frac{T - \mu(T)}{\sigma(T)}$$

- (vii) The square root of  $b_1$ , the skewness coefficient, is printed instead of  $b_1$ . The sign indicates whether the skewness is positive, i.e., long tail on the right, or negative, i.e., long tail on the left.

Storage.

To avoid any inconvenience to users, the storage in rows 1 to 43 and 51 to 60 has not been changed; even though some of the statistics have been eliminated from the automatic printing. The following statistics have been added and may be considered an addendum to page 119 of NBS Technical Note 552.

<u>Row</u>	<u>Name Of Statistic</u>	<u>Reference</u>
OTHER STATISTICS		
44	SECOND MINIMUM	Natrella (1963), 17-4
45	SECOND MAXIMUM	Natrella (1963), 17-4
46	(MEAN-MINIMUM)/S	Natrella (1963), 17-5
47	(MEAN-MAXIMUM)/S	Natrella (1963), 17-5
48	SQRT (B1)	Snedecor and Cochran (1967), 86
49	LOWER QUARTILE	Snedecor (1956), 111
50	UPPER QUARTILE	Snedecor (1956), 111
MEASURES OF DISPERSION		
61	INTER-QUARTILE RANGE	Snedecor (1956), 111
TREND STATISTICS		
62	QUADRATIC COEFFICIENT	Snedecor and Cochran (1967), 453
63	QUADRATIC SIGNIFICANCE LEVEL	Snedecor and Cochran (1967), 453
OTHER TESTS FOR NON-RANDOMNESS		
64	AUTOCORRELATION COEFFICIENT	Jenkins and Watts (1968), 182
CONFIDENCE INTERVALS		
65	LOWER CONFIDENCE LIMIT, MEDIAN	Brownlee (1965), 183
66	UPPER CONFIDENCE LIMIT, MEDIAN	Brownlee (1965), 183
TOLERANCE INTERVALS		
67	LOWER LIMIT, 50% COVERAGE	Natrella (1963) & Gardner & Hull (1966)
68	UPPER LIMIT, 50% COVERAGE	Natrella (1963) & Gardner & Hull (1966)
69	LOWER LIMIT, 95% COVERAGE	Natrella (1963) & Gardner & Hull (1966)
70	UPPER LIMIT, 95% COVERAGE	Natrella (1963) & Gardner & Hull (1966)
71	LOWER LIMIT, 99% COVERAGE	Natrella (1963) & Gardner & Hull (1966)
72	UPPER LIMIT, 99% COVERAGE	Natrella (1963) & Gardner & Hull (1966)
73	COVERAGE, 95% DISTRIBUTION FREE	Natrella (1963), 2-15

STATISTICAL ANALYSIS OF COLUMN 1

NUMBER OF MEASUREMENTS = 39. NO. OF DISTINCT MEASUREMENTS = 17

FREQUENCY DISTRIBUTION WITH 10 CLASSES OF LENGTH .24000000  
 5 3 8 3 7 7 5 0 0 1  
 FREQUENCY DISTRIBUTION OF LEAST SIGNIFICANT DIGIT (0,1,...,9)  
 6 4 6 3 4 4 4 4 3 1

MEASURES OF LOCATION .

ARITHMETIC MEAN . . . . . .41025640  
 MEDIAN . . . . . .50000000  
 MID-RANGE . . . . . .70000000  
 MID-MEAN (25 PERCENT TRIMMED MEAN) . . .42380951

MEASURES OF DISPERSION

STANDARD DEVIATION . . . . . .50668939  
 AS PERCENT OF MEAN (COEF. OF VAR.) : 123.50554  
 RANGE . . . . . .2.4000000  
 MEAN DEVIATION . . . . . .40486522  
 INTER-QUARTILE RANGE . . . . . .70000000  
 VARIANCE . . . . . .25673414

STANDARD DEVIATION OF MEAN. . . . . .081135237

TREND STATISTICS

SLOPE, SIGNIFICANCE LEVEL . . . . . -.0040080961 .585  
 QUADRATIC COEFF., SIGNIFICANCE LEVEL . .0026019551 .000

OTHER TESTS FOR NON-RANDOMNESS

NUMBER OF RUNS UP AND DOWN, Z VALUE . 23 -1.037  
 MEAN SQUARE SUCCESSIVE DIFFERENCE . . .28289473  
 MS SUCC DIFF/2(VARIANCE), Z VALUE . .55094880 -2.888  
 DEVIATIONS FROM ARITHMETIC MEAN  
 NUMBER OF - SIGNS, + SIGNS . . . . . 19 20  
 NUMBER OF RUNS, Z VALUE . . . . . 8 -4.056  
 AUTOCORRELATION COEFFICIENT . . . . . .43572843

A TWO-SIDED 95 PERCENT CONFIDENCE INTERVAL FOR THE

MEAN IS .24600670 TO .57450610  
 MEDIAN IS .20000000 TO .59999999 (DISTRIBUTION-FREE)  
 S.D. IS .41070743 TO .64762166

STATISTICAL TOLERANCE INTERVAL WITH 95 PERCENT CONFIDENCE FOR

50 PCT NORMAL COVERAGE IS -.017515361 TO .83802813  
 95 PCT NORMAL COVERAGE IS -.83254325 TO 1.6530560  
 99 PCT NORMAL COVERAGE IS -1.2228149 TO 2.0433277  
 INTERVAL FROM MIN TO MAX HAS DIST.-FREE COVERAGE 88.40

OTHER STATISTICS

MINIMUM . . . . . .- .50000000  
 SECOND MINIMUM . . . . . .- .40000000  
 MAXIMUM . . . . . .1.9000000  
 SECOND MAXIMUM . . . . . .1.1000000  
 (MEAN-MINIMUM)/STANDARD DEVIATION . .1.7964781  
 (MAXIMUM-MEAN)/STANDARD DEVIATION . .2.9401515  
 SQRT(B1), SKEWNESS COEFFICIENT . . . . .31064667  
 B2, KURTOSIS COEFFICIENT . . . . .3.3379328  
 LOWER QUARTILE . . . . . .0.  
 UPPER QUARTILE . . . . . .70000000

## OBSERVATIONS

## ORDERED OBSERVATIONS

I	X(I)	RANK	X(I) - MEAN	NO.	X(J)	X(J+1) - X(J)
1	.40000000	18.5	-.01025641	24	-.50000000	.10000000
2	.59999999	24.5	.18974359	16	-.40000000	0.00000000
3	1.00000000	35.5	.58974360	19	-.40000000	.10000000
4	1.00000000	35.5	.58974360	17	-.30000000	0.00000000
5	1.00000000	35.5	.58974360	20	-.30000000	.09999999
6	.50000000	21.0	.08974360	28	-.20000000	.10000000
7	.59999999	24.5	.18974359	22	-.10000000	0.00000000
8	.70000000	28.5	.28974359	26	-.10000000	.10000000
9	1.00000000	35.5	.58974360	15	0.	0.00000000
10	.59999999	24.5	.18974359	18	0.	.10000000
11	.20000000	14.0	-.21025641	21	.10000000	.10000000
12	1.90000000	39.0	1.4897436	11	.20000000	0.00000000
13	.20000000	14.0	-.21025641	13	.20000000	0.00000000
14	.40000000	18.5	-.01025641	23	.20000000	0.00000000
15	0.	9.5	-.41025640	27	.20000000	0.00000000
16	-.40000000	2.5	-.81025640	35	.20000000	.09999999
17	-.30000000	4.5	-.71025640	25	.30000000	.10000000
18	0.	9.5	-.41025640	1	.40000000	0.00000000
19	-.40000000	2.5	-.81025640	14	.40000000	.10000000
20	-.30000000	4.5	-.71025640	6	.50000000	0.00000000
21	.10000000	11.0	-.31025641	30	.50000000	0.00000000
22	-.10000000	7.5	-.51025640	36	.50000000	.09999999
23	.20000000	14.0	-.21025641	2	.59999999	0.00000000
24	-.50000000	1.0	-.91025640	7	.59999999	0.00000000
25	.30000000	17.0	-.11025641	10	.59999999	0.00000000
26	-.10000000	7.5	-.51025640	31	.59999999	.10000000
27	.20000000	14.0	-.21025641	8	.70000000	0.00000000
28	-.20000000	6.0	-.61025640	33	.70000000	0.00000000
29	.80000000	32.0	.38974360	34	.70000000	0.00000000
30	.50000000	21.0	.08974360	37	.70000000	.10000000
31	.59999999	24.5	.18974359	29	.80000000	0.00000000
32	.80000000	32.0	.38974360	32	.80000000	0.00000000
33	.70000000	28.5	.28974359	38	.80000000	.20000000
34	.70000000	28.5	.28974359	3	1.00000000	0.00000000
35	.20000000	14.0	-.21025641	4	1.00000000	0.00000000
36	.50000000	21.0	.08974360	5	1.00000000	0.00000000
37	.70000000	28.5	.28974359	9	1.00000000	.09999999
38	.80000000	32.0	.38974360	39	1.10000000	.80000000
39	1.10000000	38.0	.68974359	12	1.90000000	

## OBSERVATIONS

## ORDERED OBSERVATIONS

I	X(I)	RANK	X(I) -MEAN	WEIGHT	NO.	X(J)	X(J+1) -X(J)
1	.4000000	18.5	-.0102564	1.000	24	-.5000000	.1000000
2	.5999999	24.5	.1897435	1.000	16	-.4000000	0.0000000
3	1.0000000	35.5	.5897436	1.000	19	-.4000000	.1000000
4	1.0000000	35.5	.5897436	1.000	17	-.3000000	0.0000000
5	1.0000000	35.5	.5897436	1.000	20	-.3000000	.1000000
6	.5000000	21.0	.0897436	1.000	28	-.2000000	.0999999
7	.5999999	24.5	.1897435	1.000	22	-.1000000	0.0000000
8	.7000000	28.5	.2897435	1.000	26	-.1000000	.1000000
9	1.0000000	35.5	.5897436	1.000	15	0.	0.0000000
10	.5999999	24.5	.1897435	1.000	18	0.	.1000000
11	.2000000	14.0	-.2102564	1.000	21	.1000000	.1000000
12	1.9000000	39.0	1.4897436	1.000	11	.2000000	0.0000000
13	.2000000	14.0	-.2102564	1.000	13	.2000000	0.0000000
14	.4000000	18.5	-.0102564	1.000	23	.2000000	0.0000000
15	0.	9.5	-.4102564	1.000	27	.2000000	0.0000000
16	-.4000000	2.5	-.8102564	1.000	35	.2000000	.0999999
17	-.3000000	4.5	-.7102564	1.000	25	.3000000	.1000000
18	0.	9.5	-.4102564	1.000	1	.4000000	0.0000000
19	-.4000000	2.5	-.8102564	1.000	14	.4000000	.1000000
20	-.3000000	4.5	-.7102564	1.000	6	.5000000	0.0000000
21	.1000000	11.0	-.3102564	1.000	30	.5000000	0.0000000
22	-.1000000	7.5	-.5102564	1.000	36	.5000000	.0999999
23	.2000000	14.0	-.2102564	1.000	2	.5999999	0.0000000
24	-.5000000	1.0	-.9102564	1.000	7	.5999999	0.0000000
25	.3000000	17.0	-.1102564	1.000	10	.5999999	0.0000000
26	-.1000000	7.5	-.5102564	1.000	31	.5999999	.1000000
27	.2000000	14.0	-.2102564	1.000	8	.7000000	0.0000000
28	-.2000000	6.0	-.6102564	1.000	33	.7000000	0.0000000
29	.8000000	32.0	.3897436	1.000	34	.7000000	0.0000000
30	.5000000	21.0	.0897436	1.000	37	.7000000	.1000000
31	.5999999	24.5	.1897435	1.000	29	.8000000	0.0000000
32	.8000000	32.0	.3897436	1.000	32	.8000000	0.0000000
33	.7000000	28.5	.2897435	1.000	38	.8000000	.2000000
34	.7000000	28.5	.2897435	1.000	3	1.0000000	0.0000000
35	.2000000	14.0	-.2102564	1.000	4	1.0000000	0.0000000
36	.5000000	21.0	.0897436	1.000	5	1.0000000	0.0000000
37	.7000000	28.5	.2897435	1.000	9	1.0000000	.0999999
38	.8000000	32.0	.3897436	1.000	39	1.1000000	.8000000
39	1.1000000	38.0	.6897435	1.000	12	1.9000000	

STATISTICAL ANALYSIS OF PIKES PEAK

NUMBER OF MEASUREMENTS = 84, NO. OF DISTINCT MEASUREMENTS = 80

FREQUENCY DISTRIBUTION WITH 10 CLASSES OF LENGTH .014670000  
 5 25 35 8 1 0 0 4 4 2  
 FREQUENCY DISTRIBUTION OF LEAST SIGNIFICANT DIGIT (0,1,...,9)  
 11 10 13 8 10 4 5 8 12 3

MEASURES OF LOCATION

ARITHMETIC MEAN . . . . . .63734404  
 MEDIAN . . . . . .62915000  
 MID-RANGE . . . . . .66844999  
 MID-MEAN (25 PERCENT TRIMMED MEAN) . . . . .62885952

MEASURES OF DISPERSION

STANDARD DEVIATION . . . . . .032401812  
 AS PERCENT OF MEAN (COEF. OF VAR.) . . . . .5.0838808  
 RANGE . . . . . .14670000  
 MEAN DEVIATION . . . . . .021074544  
 INTER-QUARTILE RANGE . . . . . .018249989  
 VARIANCE . . . . . .0010498774

STANDARD DEVIATION OF MEAN. . . . . .0035353275

TREND STATISTICS

SLOPE, SIGNIFICANCE LEVEL . . . . . -.00024761840 .090  
 QUADRATIC COEFF., SIGNIFICANCE LEVEL . . . . . 2.1859619-06 .744

OTHER TESTS FOR NON-RANDOMNESS

NUMBER OF RUNS UP AND DOWN, Z VALUE . . . . . 47 -2.267  
 MEAN SQUARE SUCCESSIVE DIFFERENCE . . . . . .00036381001  
 MS SUCC DIFF/2(VARIANCE), Z VALUE . . . . . .17326309 -7.740

DEVIATIONS FROM ARITHMETIC MEAN

NUMBER OF - SIGNS, + SIGNS . . . . . 62 22  
 NUMBER OF RUNS, Z VALUE . . . . . 14 -5.550

AUTOCORRELATION COEFFICIENT . . . . . .82588458

A TWO-SIDED 95 PERCENT CONFIDENCE INTERVAL FOR THE

MEAN IS .63031242 TO .64437566  
 MEDIAN IS .62520000 TO .63299999 (DISTRIBUTION-FREE)  
 S.D. IS .028026469 TO .038586477

STATISTICAL TOLERANCE INTERVAL WITH 95 PERCENT CONFIDENCE FOR

50 PCT NORMAL COVERAGE IS .61210972 TO .66257809  
 95 PCT NORMAL COVERAGE IS .56402038 TO .71066743  
 99 PCT NORMAL COVERAGE IS .54098537 TO .73370244  
 INTERVAL FROM MIN TO MAX HAS DIST.-FREE COVERAGE 94.48

OTHER STATISTICS

MINIMUM . . . . . .59509999  
 SECOND MINIMUM . . . . . .59519999  
 MAXIMUM . . . . . .74180000  
 SECOND MAXIMUM . . . . . .74070000  
 (MEAN-MINIMUM)/STANDARD DEVIATION . . . . . 1.3037557  
 (MAXIMUM-MEAN)/STANDARD DEVIATION . . . . . 3.2237689  
 SQRT(B1), SKEWNESS COEFFICIENT . . . . . 1.9315909  
 B2, KURTOSIS COEFFICIENT . . . . . 5.9296504  
 LOWER QUARTILE . . . . . .61987500  
 UPPER QUARTILE . . . . . .63812499

### Some Remarks.

(1) The distribution of least significant digits can be used to assess whether the unit of measurement is actually what it is supposed to be. We may expect the distribution of least significant digits to be roughly uniform. If a distance is measured to the nearest tenth of an inch, the digits 0 to 9 after the decimal point will be roughly the same. However, if, unknown to us, the distance had been measured less precisely to the nearest one eighth of an inch with only one digit after the decimal point recorded, then the numbers 1, 4, 9 would have zero frequency. The presence of an unusually large number of zeros or fives is reason to be suspicious and to reexamine the measurement process.

(2) A 95% distribution-free confidence interval for the median cannot be computed exactly. Let  $\alpha = 1 - 0.95 = .05$  be the desired confidence coefficient. There are two approaches taken by statisticians. We can choose the confidence interval with the largest confidence coefficient,  $\alpha^*$ , such that  $\alpha^* < \alpha$ . Or, we can choose the confidence interval with confidence coefficient,  $\alpha'$  closest to  $\alpha$ . For many sample sizes the result is the same in each case. The first case is conservative. Based upon an informal poll of NBS statisticians, the second approach is being used in OMNITAB. The 95% confidence interval for the median is not printed, if the number of measurements (NRMAX) is less than 5.

### Brief

A BRIEF instruction suppresses all pages after the first page of the automatic printing. See Section C1.1.1.

### Formulas.

Let  $x_i$  be the measurements with non-zero weights and let  $y_i$  be the ordered measurements with non-zero weights.

27. The z-value for the mean square successive difference is:

$$z = (\text{MS Succ. Diff.}/2 \times \text{Variance} - 1) / \sqrt{(2n-3-b_2)/2n(n-1)}$$

44.  $y_2$

45.  $y_{k-1}$

46.  $(\bar{x}_w - y_1)/s$

47.  $(y_k - \bar{x}_w)/s$

48.  $[\sum_{i=1}^k (x_i - \bar{x}_w)^3/k] / [(k-1)s^2/k]^{3/2}$

Let  $r$  be the integral part of  $m$ . Let  $p = m - r$  be the fractional part of  $m$  and let  $q = 1 - p$ .

49.  $Q_1 = qy_r + py_{r+1}$ , where  $m = (n+1)/4$

50.  $Q_3 = py_r + qy_{r+1}$ , where  $m = 3(n+1)/4$

61.  $Q_3 - Q_1$

62. See section C4.5 of TN 552. Predictor is  $i$ .

63. See section C4.5 of TN 552. Predictor is  $i$ .

The correlation coefficient,  $r_{xy}$ , is defined on page 158 of TN 552.

64.  $r_{uv}$ , where  $u = x_1, x_2, \dots, x_{k-1}$  and

$$v = x_2, x_3, \dots, x_k$$

Let  $a$  be the value of  $c$  which makes

$$\sum_{i=0}^c \binom{k}{i} 2^{-k} \text{ closest to } 0.025$$

65.  $y_a$

66.  $y_{k+1-a}$

Gau(x) is the cumulative normal distribution defined on page 167 of TN 552.

$$\text{Let } \Pr(\text{Gau}(T_U) - \text{Gau}(T_L) = 0.50) = 0.95$$

67.  $T_L = \bar{x}_w - Ks$

68.  $T_U = \bar{x}_w + Ks$

$$\text{Let } \Pr(\text{Gau}(T_U) - \text{Gau}(T_L) = 0.95) = 0.95$$

69.  $T_L = \bar{x}_w - Ks$

70.  $T_U = \bar{x}_w + Ks$

$$\text{Let } \Pr(\text{Gau}(T_L) - \text{Gau}(T_L) = 0.99) = 0.95$$

71.  $T_L = \bar{x}_w - Ks$

72.  $T_U = \bar{x}_w + Ks$

73. The largest value of  $p$  such that

$$kp^{k-1} - (k-1)p^k \leq 0.05$$

#### 4.3\* Analysis Of Groups Of Data.

Many improvements have been made to the printing performed by a ONEWAY instruction. General improvements are:

(i) All printing has been revised to use 72 characters per line instead of 120. Hence, ONEWAY can now be used interactively with a narrow width carriage terminal or CRT display device.

(ii) All printing with scientific notation has been changed to readable printing with a constant number of significant digits with the decimal points lined up.

(iii) The restrictions on page 122 of NBS TN 552 marked by an asterisk have finally been lifted. Group numbers do not have to be (a) consecutive integers, (b) integers, or (c) positive.

Specific changes are:

(i) The addition of a two-line title at the beginning which gives the number of measurements and column heading on the first line and the number of groups and the column heading for the column containing the group numbers on the second line. If a LABEL or HEAD instruction has been used, its effect is reflected in the column heading used in this two-line title.

(ii) The printing is divided into six sections: analysis of variance, estimates, pairwise multiple comparison of means, tests for homogeneity of variances, Model II - components of variance, and a new section showing box plots. The title of each section is enclosed in a box.

(iii) When lines are printed in the analysis of variance for "SLOPE" and "ABOUT LINE" (formerly DEVS. ABOUT LINE) the SUMS OF SQUARES, MEAN SQUARES and PROB. (or significance level) are not indented, as before, because of space limitations. The H-statistic for the Kruskal-Wallis rank test is no longer printed, since the significance level contains all the information in the data.

(iv) In the printing of the estimates, the use of H and L to pinpoint the high and low values in a column has been extended to the SUM RANKS, MINIMUM, MAXIMUM, and S.D. OF MEAN columns. The blank line at the bottom of the column has been replaced by a line with dashes suitably placed. Because of space restrictions, the estimates section has been divided into two subsections. If the number of groups exceeds 14, the second subsection starts on a new page.

### Box Plots

A new section has been added, which shows box plots for each group. A box plot is a graphical display of the scrawl of a set of measurements, where the scrawl consists of the five statistics: the minimum, the lower hinge, the median, the upper hinge, and the maximum. (See Section C4.1.1.) The lower hinge and the upper hinge are essentially equivalent to the lower and upper quartiles.

An examination of a box plot enables one to get a quick assessment of the shape of the distribution of the numbers; including an assessment of location, variation and skewness. By examining all of the box plots simultaneously, the relative shapes of the distributions for the different groups can be assessed. There are several variations of box plots which are described in MCGILL et al. (1977). The simplest form is given by ONEWAY.

A box plot has a line starting with an asterisk to represent the minimum and ending with an asterisk to represent the maximum. In between, H's are appropriately placed to represent the hinges and M is appropriately placed inside to represent the median. A box is erected above and below the line from the lower hinge to the upper hinge.

In small samples, a hinge, the median, or an extreme value may coincide. In such cases, only one symbol is printed. The M, for median, takes precedence over an H, for hinge, or an \*, for an extreme value, and an \* takes precedence over an H.

Obviously, at least five measurements are needed to print a box plot. If a group has less than five measurements, an X is used to represent each measurement in the group. If two or more points coincide, a single X is printed.

The box plot section starts on a new page. A title and axis are printed at the top of the page and on any subsequent pages. Three numbers are printed above the 0's of the axis to show the scale; namely, the minimum, the midrange and the maximum of all the measurements. The group number and the number of measurements in the group are printed on the left.

Remark.

There are times when a zero group number is appropriate, as in temperatures of -20, -10, 0, 10, 20 and 30 degrees. There are times when it is desirable to delete one or more observations from an analysis by assigning a zero weight. These requirements are in conflict, unless the structure of the instruction is complicated by the addition of another argument.

Since several editing instructions were added to OMNITAB II, it has been decided to make zero a valid group number. If certain observations are to be deleted from an analysis, it can be easily done using one of the DELETE instructions described in Section C3.5.1. Suppose an analysis is needed without using the data in rows 3 and 7. Let the measurements be in column 11 and the group numbers in column 12. Pick any number, such as -99.99, which is not being used as a group number. Now the following set of instructions can be used:

```
DEFINE -99.99 into row 3 of column 12
DEFINE -99.99 into row 7 of column 12
DELETE rows with -99.99 in cols 11 and 12 and put in cols 13 and 14
ONEWAY analysis of column 13 with group numbers in column 14
```

Remark.

If a set of instructions is rerun using the latest Omnitab II version, some differences may be observed in the numerical results which will be caused by the use of more sophisticated algorithms and substantially better programming. The results will be more accurate.

Examples.

Pages 107, 108 and 109 show the printing for the instructions

```
LABEL GRAVITY G
$
SET measurements in GRAVITY G
 83 81 76 78 79 72 61 61 67 67 64 78 71 75 72 74
$
SET group numbers in column 2
 1 1 1 1 1 1 2 2 2 2 2 3 3 3 3 3
$
ONEWAY analysis of GRAVITY G with group numbers in column 2
```

The data is the same as that used to produce page 126 of NBS TN 552; except a LABEL instruction has been used to produce a column heading on the first line of the printing.

ONEWAY ANALYSIS OF 16 MEASUREMENTS IN GRAVITY G  
 CLASSIFIED INTO 3 GROUPS WITH NUMBERS IN COLUMN 2

-----  
 I ANALYSIS OF VARIANCE I  
 -----

SOURCE	D.F.	SUMS OF SQUARES	MEAN SQUARES	F RATIO	PROB.
BETWEEN GROUPS	2	565.10412	282.55206	26.082	.000
SLOPE	1	64.508909	64.508909	1.408	.257
ABOUT LINE	1	500.59521	500.59521	46.209	.000
WITHIN GROUPS	13	140.83333	10.833333		
TOTAL	15	705.93750			

KRUSKAL-WALLIS RANK TEST FOR DIFFERENCE BETWEEN MEANS  
 SIGNIFICANCE LEVEL IS APPROXIMATELY .000

-----  
 I ESTIMATES I  
 -----

GROUP	NO.	MEAN	MINIMUM	MAXIMUM	SUM RANKS
1	6	78.166666 H	72.000000 H	83.000000 H	76.0H
2	5	64.000000 L	61.000000 L	67.000000 L	15.0L
3	5	74.000000	71.000000	78.000000	45.0
TOTAL	16	72.437500	61.000000	83.000000	

GROUP	WITHIN S.D.	S.D. OF MEAN	95 PCT CONF INT FOR MEAN
1	3.8686776 H	1.5793810 H	74.106800 TO 82.226532
2	3.0000000	1.3416408	60.275069 TO 67.724931
3	2.7386128 L	1.2247449 L	70.599619 TO 77.400381
MODEL			
FIXED	3.2914029	.82285073	70.659840 TO 74.215160
RANDOM	7.2957575	4.2122076	54.313635 TO 90.561365
UNGROUPED	6.8602114	1.7150528	68.781952 TO 76.093048

-----  
I PAIRWISE MULTIPLE COMPARISON OF MEANS I  
-----

THE MEANS ARE PUT IN INCREASING ORDER IN GROUPS SEPARATED BY \*\*\*. A MEAN IS ADJUDGED NON-SIGNIFICANTLY DIFFERENT FROM ANY MEAN IN THE SAME GROUP AND SIGNIFICANTLY DIFFERENT AT THE .05 LEVEL FROM ANY MEAN IN ANOTHER GROUP. \*\*\*\*\* INDICATES ADJACENT GROUPS HAVE NO COMMON MEAN.

- NEWMAN-KEULS TECHNIQUE, HARTLEY MODIFICATION -  
(APPROXIMATE, IF THE NUMBERS OF MEASUREMENTS IN THE GROUPS DIFFER)

64.000000  
\*\*\*\*\*  
74.000000, 78.166666

- SCHEFFE TECHNIQUE -

64.000000  
\*\*\*\*\*  
74.000000, 78.166666

-----  
I TESTS FOR HOMOGENEITY OF VARIANCES I  
-----

COCHRAN'S C = MAX VARIANCE/SUM(VARIANCES) = .4756, APPROX SL = .439  
BARTLETT-BOX F = .269, SIGNIFICANCE LEVEL = .764  
MAXIMUM VARIANCE / MINIMUM VARIANCE = 1.995554

-----  
I MODEL II - COMPONENTS OF VARIANCE I  
-----

ESTIMATE OF BETWEEN COMPONENT OF VARIATION IS 51.147054

-----  
 I BOX PLOTS I  
 -----

GROUP 61.000000 72.000000 83.000000  
 NO. O-----+-----I-----+-----I-----O-----I-----+-----I-----+-----O

1 6 \*-----I H M H-----\*  
 I I  
 I I  
 -----

-----  
 I I  
 2 5 \* M \*  
 I I  
 -----

-----  
 I I  
 3 5 \*-H M H-----\*  
 I I  
 -----

Pages 111, 112, 113, 114 and 115 show the printing for an analysis of some 1976 EPA mileage data published in the Washington Post which was produced by the following set of instructions:

OMNITAB 1/12/77 USE ONEWAY WITH EPA76 DATA  
 DIMENSION worksheet to be 400 rows x 30 columns  
 WIDTH 72

LABEL COMBINED MPG,3, ENGINE SIZE,15  
 SET mileage measurements in COMBINED MPG

33	33	33	32	32	30	30	29	29	29	29	29	29	29	29	29	28	28	28	28
28	28	28	28	28	27	27	27	27	27	27	27	27	27	27	26	26	26	26	26
26	26	26	26	26	26	26	26	26	26	25	25	25	25	25	25	25	25	25	25
25	25	25	24	24	24	24	24	24	24	24	24	24	24	24	24	23	23	23	23
23	23	23	23	23	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22
22	22	22	22	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21
21	21	21	21	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	22
19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19
19	19	19	19	19	19	19	19	18	18	18	18	18	18	18	18	18	18	18	18
18	18	18	18	18	18	18	17	17	17	17	17	17	17	17	17	17	17	17	17
17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17
16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	15	15	16	16
15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13
13	13	13	13	13	13	13	13	12	12	12	12	12	12	12	12	12	12	12	12

SET measurements in ENGINE SIZE

98	85	83	85	79	129	129	97	97	91	98	85	73	91	97	97	97	97	140	97
97	97	97	97	97	140	119	119	119	140	140	140	140	85	140	140	119	140	140	140
140	140	140	140	140	140	140	97	97	97	140	140	140	119	119	119	79	200	200	200
200	140	140	119	79	79	120	120	133	133	133	133	133	133	122	258	258	258	114	140
140	140	140	140	140	231	140	140	140	226	225	225	225	107	107	231	225	225	225	140
140	231	140	122	232	258	232	231	250	225	107	107	250	171	140	140	141	171	250	231
225	225	231	152	232	232	232	114	231	231	250	250	250	250	122	168	225	107	107	250
200	171	200	250	250	250	260	225	250	250	250	250	156	156	130	130	130	134	140	
258	258	232	258	232	231	260	231	168	225	225	250	171	171	171	250	260	250	260	260
225	225	250	260	260	260	163	163	304	304	262	318	318	318	250	302	302	250	260	318
318	318	350	400	260	163	163	258	258	350	350	350	350	305	305	305	305	305	318	318
318	318	302	302	302	30	302	350	350	350	318	318	318	318	400	400	350	350	156	156
304	304	350	350	350	350	351	302	302	351	455	350	400	350	455	400	304	351	350	350
350	350	400	350	350	400	350	350	400	400	350	350	318	360	318	360	318	351	351	351
400	351	351	351	351	455	350	455	350	455	318	318	360	360	400	400	455	455	400	400
304	360	304	360	455	455	455	350	400	360	400	360	360	360	400	400	360	400	360	400
400	400	400	400	400	400	400	400	455	455	455	360	360	400	400	360	360	400	455	164
360	360	455	500	500	454	454	440	400	400	400	440	400	460	460	460	460	460	460	460
460	460	460	460	460	400	400	400	401	401	500	500	440	440	440	400	326	412	412	412

CODE column 15 with length 25.0 and put in column 11  
 MINIMUM value of column 15 put in column 12  
 MULTIPLY column 11 by 25.0 and put in column 13  
 ADD columns 12 and 13 and put in column 14  
 SUBTRACT 12.5 from column 14 and put group number in column 15  
 ONEWAY analysis of column 3 using group numbers in column 15

ONEWAY ANALYSIS OF 380 MEASUREMENTS IN COMBINED MPG  
 CLASSIFIED INTO 19 GROUPS WITH NUMBERS IN ENGINE SIZE

-----  
 I ANALYSIS OF VARIANCE I  
 -----

SOURCE	D.F.	SUMS OF SQUARES	MEAN SQUARES	F RATIO	PROB.
BETWEEN GROUPS	18	7945.3173	441.40652	157.976	.000
SLOPE	1	7047.1467	7047.1467	*****	.000
ABOUT LINE	17	898.17059	52.833564	18.909	.000
WITHIN GROUPS	361	1008.6827	2.7941350		
TOTAL	379	8954.0000			

KRUSKAL-WALLIS RANK TEST FOR DIFFERENCE BETWEEN MEANS  
 SIGNIFICANCE LEVEL IS APPROXIMATELY .000

-----  
 I ESTIMATES I  
 -----

GROUP	NO.	MEAN	MINIMUM	MAXIMUM	SUM RANKS
42.5	1	17.000000	17.000000	17.000000	157.0
67.5	5	26.800000	24.000000	32.000000	1692.5
92.5	24	28.791667	H 26.000000	H 33.000000	H 8701.5
117.5	23	23.869565	20.000000	30.000000	6995.0
142.5	50	24.000000	20.000000	28.000000	15480.5H
167.5	17	18.823529	14.000000	21.000000	3458.5
192.5	6	23.333333	20.000000	25.000000	1771.0
217.5	16	20.937500	19.000000	22.000000	4137.0
242.5	39	20.051282	18.000000	22.000000	9264.0
267.5	20	19.400000	17.000000	23.000000	4331.5
292.5	15	16.466666	14.000000	18.000000	2108.5
317.5	25	16.640000	12.000000	L 18.000000	3653.0
342.5	39	15.769231	14.000000	18.000000	4798.5
367.5	18	14.111111	13.000000	15.000000	1185.0
392.5	41	14.243902	12.000000	L 18.000000	2907.0
417.5	3	12.000000	L 12.000000	L 12.000000	L 19.5L
442.5	7	12.571429	12.000000	L 13.000000	125.5
467.5	27	13.851852	13.000000	16.000000	1538.5
492.5	4	12.500000	12.000000	L 13.000000	66.0
TOTAL	380	19.000000	12.000000	33.000000	

GROUP	WITHIN S.D.	S.D. OF MEAN	95 PCT CONF INT FOR MEAN
42.5	*** ESTIMATES ARE NOT AVAILABLE BECAUSE SAMPLE SIZE IS 1 ***		
67.5	3.5637060 H	1.5937378 H	22.375147 TO 31.224853
92.5	2.0847097	.42553958	27.911371 TO 29.671962
117.5	3.0196982	.62965059	22.563750 TO 25.175380
142.5	2.2497165	.31815796	23.360638 TO 24.639362
167.5	1.7042335	.41333732	17.947294 TO 19.699765
192.5	2.5819889	1.0540925	20.623744 TO 26.042923
217.5	1.2893797	.32234492	20.250438 TO 21.624562
242.5	.97194102	.15563512	19.736215 TO 20.366349
267.5	1.7888544	.39999999	18.562790 TO 20.237209
292.5	1.3557637	.35005668	15.715870 TO 17.217463
317.5	1.4106736	.28213472	16.057703 TO 17.222297
342.5	.93080440	.14904799	15.467499 TO 16.070962
367.5	.58298308	.13741043	13.821200 TO 14.401022
392.5	1.2996247	.20296728	13.833690 TO 14.654114
417.5	0. L	0. L	12.000000 TO 12.000000
442.5	.53452248	.20203051	12.077080 TO 13.065777
467.5	.98853831	.19024429	13.460799 TO 14.242904
492.5	.57735027	.28867513	11.581320 TO 13.418680
MODEL	-----		
FIXED	1.6715666	.085749559	18.831369 TO 19.168631
RANDOM	5.0623025	1.1613718	16.560049 TO 21.439951
UNGROUPED	4.8605894	.24934298	18.509731 TO 19.490269

-----  
 I PAIRWISE MULTIPLE COMPARISION OF MEANS I  
 -----

THE MEANS ARE PUT IN INCREASING ORDER IN GROUPS SEPARATED BY \*\*\*. A MEAN IS ADJUDGED NON-SIGNIFICANTLY DIFFERENT FROM ANY MEAN IN THE SAME GROUP AND SIGNIFICANTLY DIFFERENT AT THE .05 LEVEL FROM ANY MEAN IN ANOTHER GROUP. \*\*\*\*\* INDICATES ADJACENT GROUPS HAVE NO COMMON MEAN.

- NEWMAN-KEULS TECHNIQUE, HARTLEY MODIFICATION -  
 (APPROXIMATE, IF THE NUMBERS OF MEASUREMENTS IN THE GROUPS DIFFER)

12.000000, 12.500000, 12.571429, 13.851852, 14.111111,  
 15.769231, 16.466666, 16.640000, 17.000000  
 \*\*\*\*\*  
 18.823529, 19.400000  
 \*\*\*  
 19.400000, 20.051282  
 \*\*\*  
 20.051282, 20.937500  
 \*\*\*\*\*  
 23.333333, 23.869565, 24.000000  
 \*\*\*\*\*  
 26.800000  
 \*\*\*\*\*  
 28.791667

- SCHEFFE TECHNIQUE -

12.000000, 12.500000, 12.571429, 13.851852, 14.111111,  
 15.769231, 16.466666, 16.640000, 17.000000, 18.823529,  
 20.051282, 20.937500  
 \*\*\*  
 18.823529, 19.400000, 20.051282, 20.937500, 23.333333,  
 \*\*\*  
 20.937500, 23.333333, 23.869565, 24.000000  
 \*\*\*  
 23.869565, 24.000000, 26.800000, 28.791667

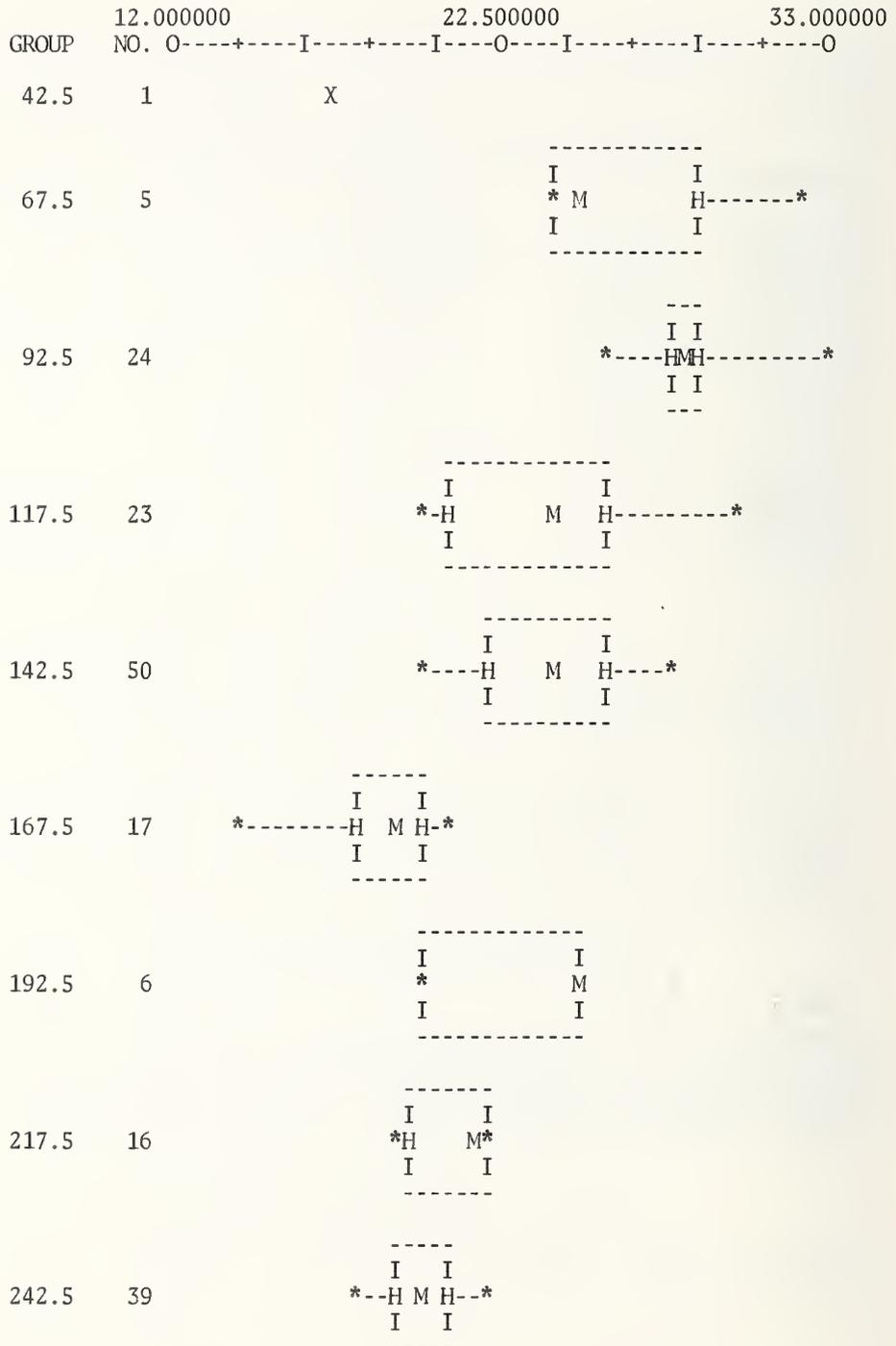
-----  
 I TESTS FOR HOMOGENEITY OF VARIANCES I  
 -----

COCHRAN'S C = MAX VARIANCE/SUM(VARIANCES) = .2312, APPROX SL = .000  
 BARTLETT-BOX F = 7.660, SIGNIFICANCE LEVEL = .000  
 MAXIMUM VARIANCE / MINIMUM VARIANCE = 0.  
 ASSUMING HETROGENEOUS VARIANCE, APPROX. BETWEEN MEANS F-TEST SL = .000

-----  
 I MODEL II - COMPONENTS OF VARIANCE I  
 -----

ESTIMATE OF BETWEEN COMPONENT OF VARIATION IS 22.542293

I BOX PLOTS I



12.000000 22.500000 33.000000  
 GROUP NO. O-----+-----I-----+-----I-----O-----I-----+-----I-----+-----O

267.5 20  
 -----  
 I I  
 \*-H MH-----\*  
 I I  
 -----

292.5 15  
 -----  
 I I  
 \*-----H M \*  
 I I  
 -----

317.5 25  
 -----  
 I  
 \*-----M-\*  
 I  
 -----

342.5 39  
 -----  
 I I  
 \*-H M H-\*  
 I I  
 -----

367.5 18  
 -----  
 I  
 \*-M-\*  
 I  
 -----

392.5 41  
 -----  
 I I  
 \*-----M H-----\*  
 I I  
 -----

417.5 3 X

442.5 7  
 -----  
 I I  
 \* M  
 I I  
 -----

467.5 27  
 -----  
 I I  
 \* M H--\*  
 I I  
 -----

492.5 4 X X

#### 4.4\* Analysis Of Two-way Table.

Some minor changes have been made in the printing produced by a TWOWAY instruction. Some examples are given in section C4.4.1.

##### 4.4.1 Two-way Fixed Effects Design With Replication.

A TWOWAY instruction can be used to analyze data from a row by column classification with one or no measurements in each cell. If there are more than one measurement in any cell, a TWOWAY instruction cannot be used directly. However, it can be used indirectly, as is explained below. The method outlined is practical, but would be improved substantially by a specific instruction which may be added in the future.

The analysis of variance table for a two-way classification with replication is:

<u>Source</u>	<u>d.f.</u>	<u>Sums of Squares</u>	<u>Mean Squares</u>	<u>F Ratio</u>	<u>Significance Level</u>
Rows	DR	SR	MR	FR	PR
Columns	DC	SC	MC	FC	PC
Row x Column Interaction	DI	SI	MI	FI	PI
Residual	DE	SE	ME		
Total	DT	ST			

The above table can be easily constructed by hand after using a set of OMNITAB instructions to obtain the necessary sums of squares and mean squares.

The degrees of freedom are calculated as follows:

$$DR = r - 1, \text{ where } r = \text{number of rows}$$

$$DC = c - 1, \text{ where } c = \text{number of columns}$$

$$DI = (r - 1)(c - 1)$$

$$DE = \text{SUM}(n_i - 1), \text{ where } n_i \text{ is the number of measurements in the } i\text{th cell and } i = 1, 2, \dots, rc$$

$$DT = \text{SUM}(n_i) - 1$$

The sums of squares can be calculated in a number of ways. The following procedure is suggested. Enter the measurements into the worksheet using a SET instruction. Using another SET instruction, enter group numbers for a ONEWAY analysis. Set the group number equal to 1 for the measurements in the first row of the first column of the rxc table, 2 for the measurements in the first row of the second column, etc., until numbers for the first row have been assigned. Continue, using consecutive integers, for the second row, third row, etc. Do a ONEWAY analysis using the form of ONEWAY with six arguments for storing the group number, number of measurements in each group, the means and the standard deviations. Then do a TWOWAY analysis using the means from ONEWAY as measurements and the numbers in the groups as the weights.

The sums of squares for the analysis of variance are obtained as follows:

SR = row sum of squares from TWOWAY  
SC = column sum of squares from TWOWAY  
SI = residual sum of squares from TWOWAY  
SE = within sum of squares from ONEWAY  
ST = total sum of squares from ONEWAY

Note, the row and column sums of squares from TWOWAY should be chosen from the line showing a value for F PROB. Note that  $SR + SC + SI + SE$  will not equal  $ST$ , unless there is an equal number of measurements in each cell or the cell frequencies are proportional.

The mean squares are obtained, as usual, by dividing the sums of squares by the corresponding degrees of freedom. For a fixed effects model, the F-ratios  $FR$ ,  $FC$  and  $FI$  are obtained by dividing  $MR$ ,  $MC$  and  $MI$  by  $ME$ .

An F-test should be done to assess whether there is a significant row by column interaction. If there is a significant interaction, it makes little sense to test for row or column effects. Instead, a graphical study of the means should be done to determine the reason for the row by column interaction. A significant interaction may be caused by an outlier or an inappropriate scale for the response measurements. If an outlier is present, an attempt should be made to determine the cause. A reanalysis without using the outlier may be appropriate. If an outlier is not present, a transformation (such as log or square root) of the response variable may remove the interaction. Probability plots of residuals may be helpful.

Note, the BETWEEN GROUPS sum of squares from the ONEWAY analysis should equal the TOTAL sum of squares from the TWOWAY analysis.

Example.

The following data has been drawn from a medical thermometry experiment. The data consisted of measurements of bead thermistor drift rates for five manufacturers and three resistances: low, medium and high. A set of OMNITAB instructions to analyze the data is:

```

OMNITAB 2/16/77 ANALYSIS OF MEDICAL THERMOMETRY DRIFT RATES
WIDTH 80
SET DRIFT RATE IN COLUMN 1
-.00146271 .02112775 .00282039 -.02015703 -.00576177 .00087166 .00195146
.00486606 .00023100 -.03231597 -.05563992 .01034201 -.00423480 -.00661010
-.01153926 .00618470 .00389811 .01307638 .00128444 .00779145 .01144474
.00788975 -.02578416 .00356302 .01287440 -.05037265 -.00086367 -.00673810
.00921573 .00604309 .00219059 .01001163 .01505623 .00875096 .00931758
-.00713043 .02330194 -.00255854 .02775204 .01250641 .00207295 -.00672891
.00517192 -.00761442 .03754005 .00952288 .00981737 .00480323 .02078291
.01174765 -.22716029 -.01801598 .01118366 .00146129 .02934898 .01127789
-.05363813 .00084837
SET ROW-COLUMN CELL NUMBER IN COLUMN 2
1 1 1 1 1 2 2 2 2 3 3 4 4 4
4 4 5 5 5 5 5 6 6 6 6 6 7 7
7 7 8 8 8 8 8 9 9 9 9 9 10 10
10 10 10 11 11 11 11 11 12 12 12 15 15 15 15 15
ONEWAY ANALYSIS OF COLUMN 1 WITH IDENTIFICATION IN COLUMN 2 STORE 11 12 13 14
DEFINE 13,12 INTO 15,12
DEFINE 13,13 INTO 15,13
ADEFINE 13,12 SIZE 2X2 EQUAL TO 0.0
RESET 15
TWOWAY ANALYSIS FOR 5X3 TABLE OF MEANS IN COL 13 STORE 21 WEIGHTS 12

```

The results are shown on pages 118 to 123.

ONEWAY ANALYSIS OF 58 MEASUREMENTS IN COLUMN 1  
 CLASSIFIED INTO 13 GROUPS WITH NUMBERS IN COLUMN 2

-----  
 I ANALYSIS OF VARIANCE I  
 -----

SOURCE	D.F.	SUMS OF SQUARES	MEAN SQUARES	F RATIO	PROB.
BETWEEN GROUPS	12	.024407991	.0020339992	2.041	.042
SLOPE	1	.00022141700	.00022141700	.180	.674
ABOUT LINE	11	.024186574	.0021987795	2.207	.027
WITHIN GROUPS	45	.044841114	.00099646919		
TOTAL	57	.069249106			

KRUSKAL-WALLIS RANK TEST FOR DIFFERENCE BETWEEN MEANS  
 SIGNIFICANCE LEVEL IS APPROXIMATELY .205

-----  
 I ESTIMATES I  
 -----

GROUP	NO.	MEAN	MINIMUM	MAXIMUM	SUM RANKS
1	5	-.00068667396	-.020157030	.021127750	122.0
2	4	.0019800450	.00023100000	.0048660600	99.0
3	2	-.043977945	-.055639920	-.032315970	L 7.0L
4	5	-.0011714900	-.011539260	.010342010	118.0
5	5	.0074990239	.0012844400	.013076380	187.0
6	5	-.010365928	-.050372650	.012874400	126.0
7	4	.0019142625	-.0067381000	.0092157299	104.0
8	5	.0090653979	.0021905900	.015056230	200.0
9	5	.010774284	-.0071304300	.027752040	188.0
10	5	.0060883179	-.0076144200	.037540050	H 140.0
11	5	.011334808	H .0048032300	H .020782910	215.0H
12	3	-.077997536	L -.22716029	L .011183660	54.0
15	5	-.0021403202	-.053638130	.029348980	151.0
TOTAL	58	-.0026617961	-.22716029	.037540050	

GROUP	WITHIN S.D.	S.D. OF MEAN	95 PCT CONF INT FOR MEAN
1	.014946757	.0066843929	-.019245222 TO .017871874
2	.0020508190 L	.0010254095 L	-.0012832181 TO .0052433081
3	.016492523	.011661975	-.19215732 TO .10420144
4	.0091258864	.0040812204	-.012502590 TO .010159610
5	.0049585463	.0022175293	.0013422756 TO .013655772
6	.026952077	.012053335	-.043830808 TO .023098952
7	.0071400029	.0035700015	-.0094469101 TO .013275435
8	.0045883242	.0020519609	.0033683336 TO .014762462
9	.015382860	.0068794243	-.0083257495 TO .029874317
10	.018427023	.0082408153	-.016791482 TO .028968117
11	.0058669472	.0026237785	.0040501492 TO .018619466
12	.13000115 H	.075056201 H	-.40094185 TO .24494678
15	.031006260	.013866421	-.040639051 TO .036358411
MODEL			
FIXED	.031566900	.0041449370	-.011010128 TO .0056865356
RANDOM	.026115926	.0072432547	-.018443490 TO .013119898
UNGROUPED	.034855367	.0045767339	-.011826547 TO .0065029552

-----  
 I PAIRWISE MULTIPLE COMPARISION OF MEANS I  
 -----

THE MEANS ARE PUT IN INCREASING ORDER IN GROUPS SEPARATED BY \*\*\*. A MEAN IS ADJUDGED NON-SIGNIFICANTLY DIFFERENT FROM ANY MEAN IN THE SAME GROUP AND SIGNIFICANTLY DIFFERENT AT THE .05 LEVEL FROM ANY MEAN IN ANOTHER GROUP. \*\*\*\*\* INDICATES ADJACENT GROUPS HAVE NO COMMON MEAN.

- NEWMAN-KEULS TECHNIQUE, HARTLEY MODIFICATION -  
 (APPROXIMATE, IF THE NUMBERS OF MEASUREMENTS IN THE GROUPS DIFFER)

-.077997536, -.043977945

\*\*\*

-.043977945, -.010365928, -.0021403202, -.0011714900, -.00068667396,  
 .0019800450, .0060883179, .0074990239, .0090653979, .010774284,

- SCHEFFE TECHNIQUE -

-.077997536, -.043977945, -.010365928, -.0021403202, -.0011714900,  
 .0019142625, .0019800450, .0060883179, .0074990239, .0090653979,  
 .011334808

-----  
 I TESTS FOR HOMOGENEITY OF VARIANCES I  
 -----

COCHRAN'S C = MAX VARIANCE/SUM(VARIANCES) = .8502, APPROX SL = .000  
 BARTLETT-BOX F = 7.386, SIGNIFICANCE LEVEL = .000  
 MAXIMUM VARIANCE / MINIMUM VARIANCE = 4018.2760  
 ASSUMING HETROGENEOUS VARIANCE, APPROX. BETWEEN MEANS F-TEST SL = .000

-----  
 I MODEL II - COMPONENTS OF VARIANCE I  
 -----

ESTIMATE OF BETWEEN COMPONENT OF VARIATION IS .00023339397

-----  
 I BOX PLOTS I  
 -----

GROUP	NO.	-.22716029	-.094810121	.037540050
		O-----+-----I-----+-----I-----O-----I-----+-----I-----+-----O		
1	5			-- II *--HM---* II --
2	4			XX
3	2		X X	
4	5			--- I I *M H* I I ---
5	5			-- II *M* II --
6	5			----- I I *---H M* I I -----
7	4			XXXX
8	5			- I *-M* I -
9	5			----- I I * M H* I I -----
10	5			--- I I *HMH-----* I I ---



ANALYSIS OF VARIANCE FOR TWO-WAY 5 X 3 TABLE

SOURCE	D.F.	SUMS OF SQUARES	MEAN SQUARES	F RATIO	F PROB.
BETWEEN ROWS	4	.0054699485	.0013674871	.646	.650
BETWEEN COLS	2	.0062415992	.0031207996		
RESIDUALS	6	.012696442	.0021160737		
TOTAL	12	.024407992			
BETWEEN ROWS	4	.0027475457	.00068688644		
BETWEEN COLS	2	.0089640018	.0044820009	2.118	.201
RESIDUALS	6	.012696442	.0021160737		
TOTAL	12	.024407992			

A WEIGHTED LEAST SQUARES ANALYSIS USING  
 13 NON-ZERO WEIGHTS AND 2 ZERO WEIGHTS IN COLUMN 12

COEFFICIENT		ESTIMATE	STD. DEV.
GRAND MEAN		-.00042717298	.0066603256
ROW	1	-.011283253	.013046438
ROW	2	-.00091895890	.011357944
ROW	3	.0085589782	.011603147
ROW	4	-.013821500	.012155542
ROW	5	.017464733	.019091041
COLUMN	1	.0069898283	.0088861858
COLUMN	2	.012188053	.0088157885
COLUMN	3	-.019177881	.0094682242
RESIDUAL			.046000801

5 X 3 TABLE OF RESIDUALS, STANDARDIZED BY  
DIVIDING EACH RESIDUAL BY ITS STANDARD DEVIATION.

COLUMN ROW	1	2	3
1	.30	.09	-.48
2	-.48	-.23	.73
3	-.78	-.80	1.59
4	.96	.96	-2.17
5	.00	.00	.00



An analysis of variance for these data is:

<u>Source</u>	<u>d.f.</u>	<u>Sums of Squares</u>	<u>Mean Squares</u>	<u>F Ratio</u>
Manufacturers (Rows)	4	.0054699485	.0013674871	
Resistance (Cols)	2	.0062415992	.0031207996	
Interaction	6	.012696442	.0021160737	2.12
Residual	45	.044841114	.00099646919	
Total	57	.069249106		

Note, that no measurements were available for the fourth manufacturer for low and medium resistance. Hence, two degrees of freedom are lost for estimating the manufacturers x resistance interaction.

After an examination of the ONEWAY analysis, in this particular example, one might decide not to proceed with a TWOWAY analysis. It is clear from the box plots that one of the high resistance measurements for the fourth manufacturer is substantially smaller than any of the other measurements. An attempt should be made to determine the cause of this low measurement before proceeding with any further analysis.

The interaction F-ratio is almost significant at the 0.05 level. No sophisticated analysis is needed, as the interaction is undoubtedly due to the outlier detected by an examination of the box plots.

It is a good idea to examine plots of the residuals. The instruction:

FIT 13, 12, 7, 21\*\*\*27

will give the four plots of residuals shown on page 124.

#### 4.5\* Regression.

Minor changes have been made in the printing of the FIT and POLYFIT instructions for batch use. Columns three and four of the analysis of variance have been deleted (CUM. MS REDUCTION and D.F.). Titles have been changed slightly. In some cases, columns titled COLUMN are now titled INDEP. VAR. and the numbers below replaced by COLUMN xxxx, where xxxx is the column number, or the heading if a LABEL or HEAD instruction has been used. Pages 127 to 130 show the results of using the set of instructions at the bottom of page 139 of NBS TN 552 and, in effect, replace pages 140 to 143 of NBS TN 552.

The printing has been substantially revised for interactive use or use after a WIDTH instruction with an argument less than 120 so that all printing uses 72 characters per line. Minor changes have been made in the first page. Two pages are used instead of one for the four plots of standardized residuals. The analysis of variance is split into two sections and one is printed below the other. At the end the printing for the fit omitting the last independent variable (term) appears below, instead of on the right of, the printing of the coefficients and their standard deviations.

Pages 131 to 135 show the printing for the example on page 139 of NBS TN 552 using:

```
WIDTH 80
POLYFIT COL 1 WTS 1.0, DEGREE 1 WITH X IN COL 3
```

instead of the instructions

```
DEFINE 1.0 INTO COLUMN 2
FIT COL 1 WTS 1.0, 2 VEC'S IN COLS 2 and 3
```

Note, on the first page of printing (page 131) the weights are not printed on the right. Values of the independent variable are printed on the right where the weights are normally printed. However, if the second argument is a column number or a constant not equal to one, the weights are printed instead of values of the independent variable.

### Brief

If a BRIEF instruction is used:

- (a) the first page is not printed.
- (b) the last two plots of standardized residuals are not printed (lower half of page with normal printing).
- (c) the variance-covariance matrix is not printed and the second half of the analysis of variance is not printed.
- (d) the printing for the fit omitting the last term is omitted.

Pages 136 to 138 show the results of using the instructions

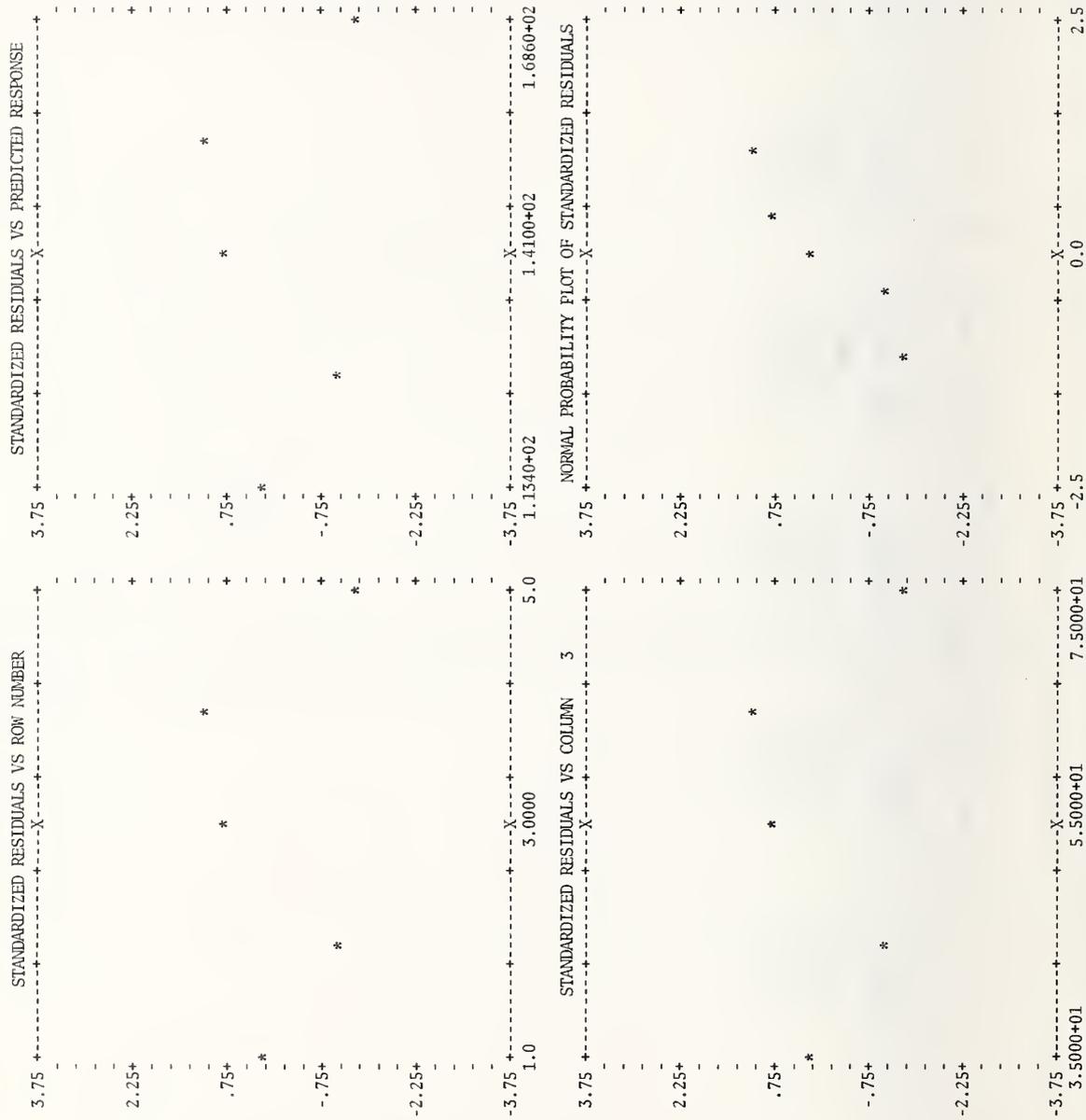
```
BRIEF
WIDTH 80
POLYFIT COL 1 WTS 1.0, DEGREE 1 WITH X IN COL 3
```

with the example, already discussed, on page 139 of NBS TN 552.

After we have had more experience, the action taken by BRIEF may be modified. An optional form may be added with an argument, which works something like LIST (n) to selectively delete parts of the printing.

LEAST SQUARES FIT OF RESPONSE, COLUMN 1,  
 AS A LINEAR FUNCTION OF 2 INDEPENDENT VARIABLES IN COLUMNS 2, 3  
 USING 5 NON-ZERO WEIGHTS = 1.0000000

ROW	INDEP VAR. COLUMN 3	RESPONSE COLUMN 1	PREDICTED RESPONSE	STD. DEV. OF PRED. RESPONSE	RESIDUALS	STD. RES.	WEIGHTS
1	35.000000	114.00000	113.40000	2.5139609	.59999935	.29	1.000
2	45.000000	124.00000	127.20000	1.7776388	-3.2000003	-1.18	1.000
3	55.000000	143.00000	141.00000	1.4514360	2.0000000	.69	1.000
4	65.000000	158.00000	154.80000	1.7776388	3.2000003	1.18	1.000
5	75.000000	166.00000	168.60000	2.5139609	-2.5999994	-1.27	1.000



LEAST SQUARES FIT OF RESPONSE, COLUMN 1,  
 AS A LINEAR FUNCTION OF 2 INDEPENDENT VARIABLES IN COLUMNS 2, 3  
 USING 5 NON-ZERO WEIGHTS = 1.0000000

VARIANCE-COVARIANCE MATRIX OF THE ESTIMATED COEFFICIENTS

COLUMN	2	3
2	33.969998	
3	-.57933331	.010533333

ANALYSIS OF VARIANCE  
 -DEPENDENT ON ORDER INDEPENDENT VARIABLES ARE ENTERED, UNLESS VECTORS ARE ORTHOGONAL-

INDEP VAR.	SS=RED. DUE TO COEF.	CUM. RESIDUAL MS	D.F.	F (COEF=0)	P (F)	F (COEFS=0)	P (F)
COLUMN 2	99404.996	483.99998	4	9437.184	.000	4808.990	.000
COLUMN 3	1904.3999	10.533333	3	180.797	.001	180.797	.001
RESIDUAL	31.599998		3				
TOTAL	101341.00		5				

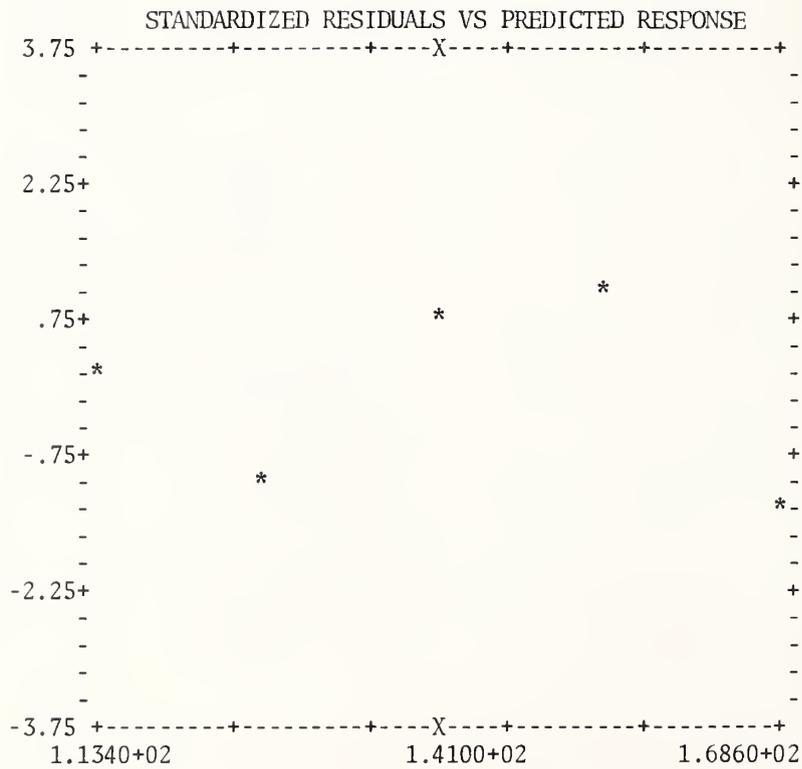
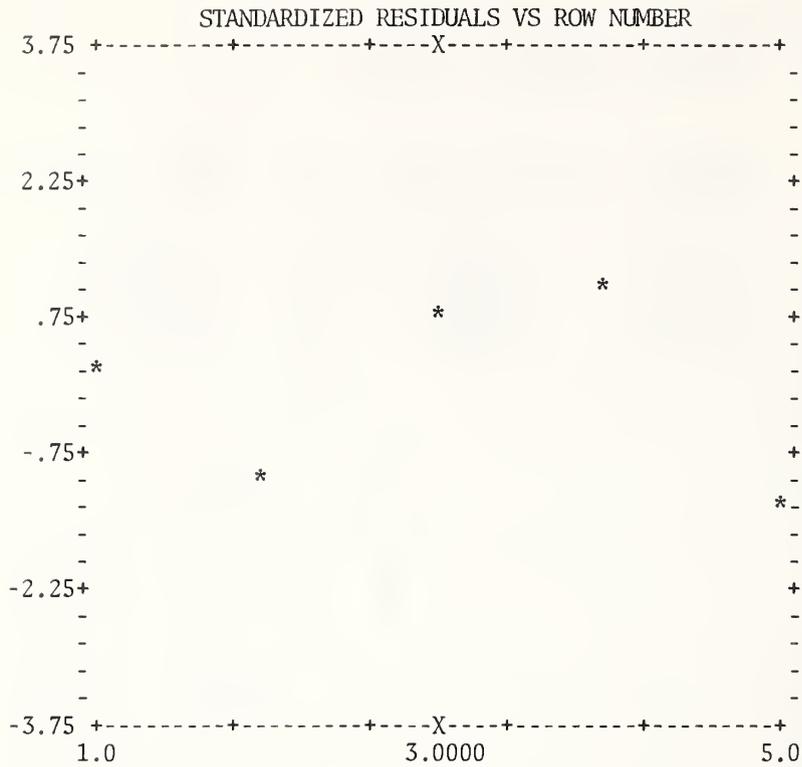
LEAST SQUARES FIT OF RESPONSE, COLUMN 1,  
 AS A LINEAR FUNCTION OF 2 INDEPENDENT VARIABLES IN COLUMNS 2, 3  
 USING 5 NON-ZERO WEIGHTS = 1.0000000

ESTIMATES FROM LEAST SQUARES FIT		FIT OMITTING LAST INDEPENDENT VAR.					
INDEP VAR.	COEFFICIENT	S.D. OF COEFF.	RATIO	ACCURACY*	COEFFICIENT	S.D. OF COEFF.	RATIO
COLUMN 2	65.100002	5.8283787	11.17	8.00	141.00000	9.8386990	14.35
COLUMN 3	1.3800000	.10263205	13.45	8.00			
RESIDUAL STANDARD DEVIATION =		3.2455097		22.000000			
BASED ON DEGREES OF FREEDOM		5 - 2 = 3		5 - 1 = 4			

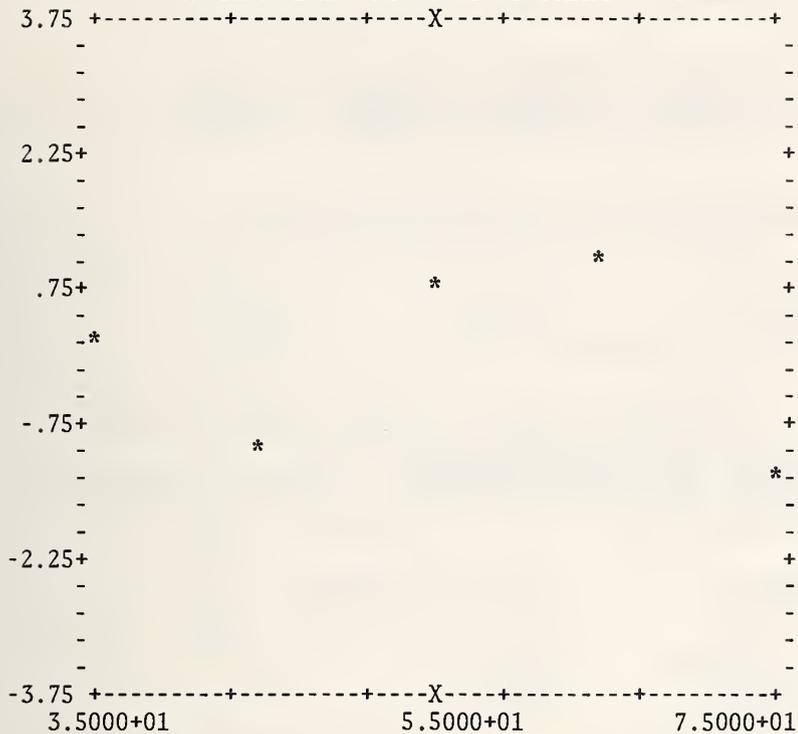
\* THE NUMBER OF CORRECTLY COMPUTED DIGITS IN EACH COEFFICIENT USUALLY DIFFERS BY LESS THAN 1 FROM THE NUMBER GIVEN HERE

LEAST SQUARES FIT OF RESPONSE, COLUMN 1,  
AS A POLYNOMIAL OF DEGREE 1. INDEPENDENT VARIABLE IS COLUMN 3  
USING 5 NON-ZERO WEIGHTS = 1.0000000

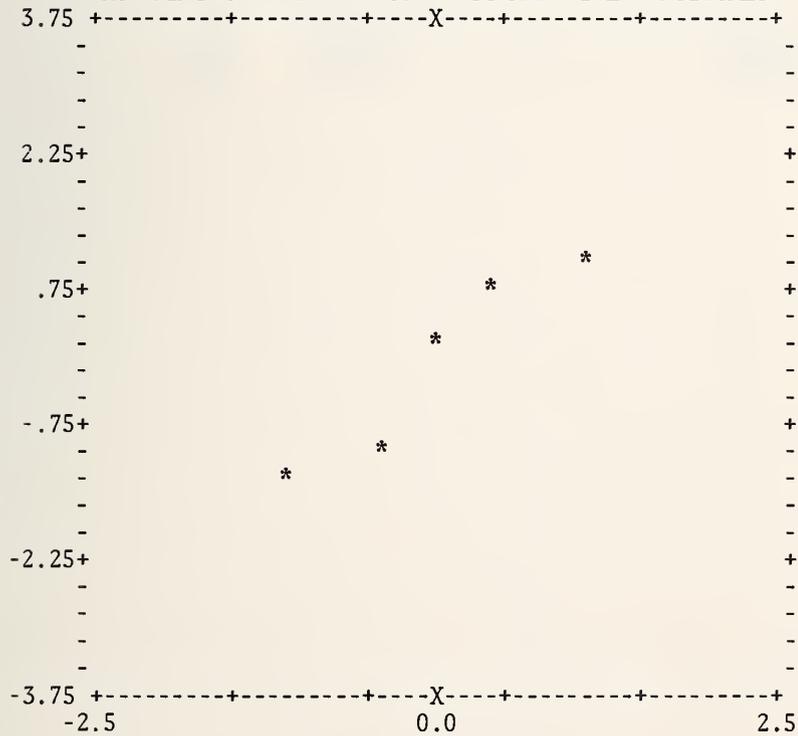
ROW	RESPONSE COLUMN 1	PREDICTED RESPONSE	STD. DEV. OF PRED. RESPONSE	STD. RES.	INDEP VAR. COLUMN 3
1	114.00000	113.40000	2.5139609	.29	35.000000
2	124.00000	127.20000	1.7776388	-1.18	45.000000
3	143.00000	141.00000	1.4514360	.69	55.000000
4	158.00000	154.80000	1.7776388	1.18	65.000000
5	166.00000	168.60000	2.5139609	-1.27	75.000000



STANDARDIZED RESIDUALS VS COLUMN 3



NORMAL PROBABILITY PLOT OF STANDARDIZED RESIDUALS



LEAST SQUARES FIT OF RESPONSE, COLUMN 1,  
 AS A POLYNOMIAL OF DEGREE 1. INDEPENDENT VARIABLE IS COLUMN 3  
 USING 5 NON-ZERO WEIGHTS = 1.0000000

VARIANCE-COVARIANCE MATRIX OF THE ESTIMATED COEFFICIENTS

TERM	0	1
0	33.969998	
1	-.57933331	.010533333

ANALYSIS OF VARIANCE  
 -DEPENDENT ON ORDER INDEPENDENT VARIABLES ENTERED,  
 UNLESS VECTORS ARE ORTHOGONAL-

TERM	SS=RED. DUE TO COEF.	F(COEF=0)	P(F)
0	99404.996	9437.184	.000
1	1904.3999	180.797	.001
RESIDUAL	31.599998	D.F. = 3	
TOTAL	101341.00	D.F. = 5	

TERM	CUM. RESIDUAL MS	D.F.	F(COEF=0)	P(F)
0	483.99998	4	4808.990	.000
1	10.533333	3	180.797	.001

LEAST SQUARES FIT OF RESPONSE, COLUMN 1,  
 AS A POLYNOMIAL OF DEGREE 1. INDEPENDENT VARIABLE IS COLUMN 3  
 USING 5 NON-ZERO WEIGHTS = 1.0000000

ESTIMATES FROM LEAST SQUARES FIT

TERM	COEFFICIENT	S.D. OF COEFF.	RATIO	ACCURACY*
0	65.100002	5.8283787	11.17	8.00
1	1.3800000	.10263203	13.45	8.00

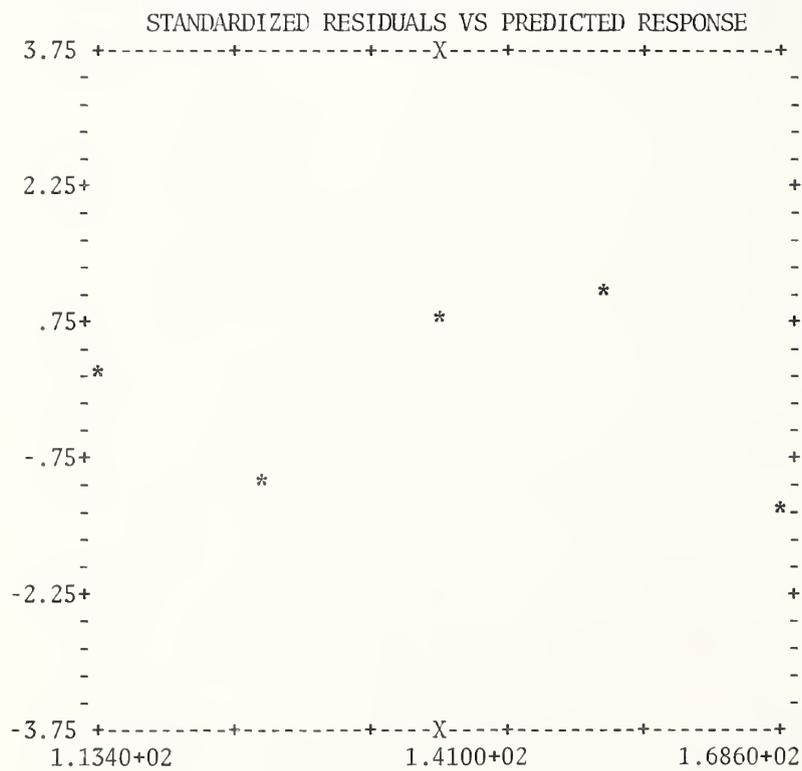
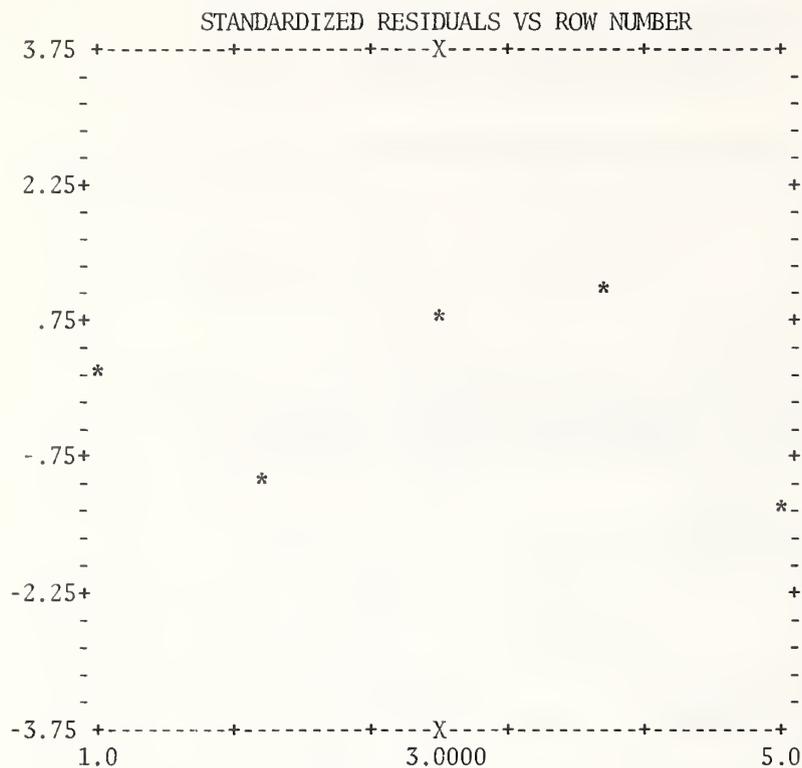
RESIDUAL STANDARD DEVIATION = 3.2455097  
 BASED ON DEGREES OF FREEDOM 5 - 2 = 3

\* THE NUMBER OF CORRECTLY COMPUTED DIGITS IN EACH COEFFICIENT  
 USUALLY DIFFERS BY LESS THAN 1 FROM THE NUMBER GIVEN HERE.

FIT OMITTING LAST TERM

TERM	COEFFICIENT	S.D. OF COEFF.	RATIO
0	141.00000	9.8386990	14.33

RESIDUAL STANDARD DEVIATION = 22.000000  
 BASED ON DEGREES OF FREEDOM 5 - 1 = 4



LEAST SQUARES FIT OF RESPONSE, COLUMN 1,  
 AS A POLYNOMIAL OF DEGREE 1. INDEPENDENT VARIABLE IS COLUMN 3  
 USING 5 NON-ZERO WEIGHTS = 1.0000000

ANALYSIS OF VARIANCE  
 -DEPENDENT ON ORDER INDEPENDENT VARIABLES ENTERED,  
 UNLESS VECTORS ARE ORTHOGONAL-

TERM	SS=RED. DUE TO COEF.	F(COEF=0)	P(F)
0	99404.996	9437.184	.000
1	1904.3999	180.797	.001
RESIDUAL	31.599998	D.F. = 3	
TOTAL	101341.00	D.F. = 5	

LEAST SQUARES FIT OF RESPONSE, COLUMN 1,  
 AS A POLYNOMIAL OF DEGREE 1. INDEPENDENT VARIABLE IS COLUMN 3  
 USING 5 NON-ZERO WEIGHTS = 1.000000

## ESTIMATES FROM LEAST SQUARES FIT

TERM	COEFFICIENT	S.D. OF COEFF.	RATIO	ACCURACY*
0	65.100002	5.8283787	11.17	8.00
1	1.3800000	.10263203	13.45	8.00

RESIDUAL STANDARD DEVIATION = 3.2455097  
 BASED ON DEGREES OF FREEDOM 5 - 2 = 3

\* THE NUMBER OF CORRECTLY COMPUTED DIGITS IN EACH COEFFICIENT  
 USUALLY DIFFERS BY LESS THAN 1 FROM THE NUMBER GIVEN HERE.

#### 4.5.1 Selection Of Variables In Linear Regression.

Often an experimenter starts with (k) predictors (independent variables) in a linear regression problem, but wants to use a model with only (p) predictors, where (p) is often much smaller than (k). The objective is to obtain a parsimonious model which describes the data satisfactorily and is subject to little or no bias.

The problem is often referred to as one of selecting the "best" subset of the (k) predictor variables. It is not a straightforward problem. The meaning of "best" may vary from one problem to another and may not be clear in any one problem. Hocking (1976) gives a comprehensive review of the literature.

Stepwise regression is often used in variable subset selection. The method has a number of disadvantages. It yields a single answer, which may not be the "best." Both forward selection and backward selection procedures can be used. They do not necessarily give the same answer and when they do, the answer is not necessarily the "best" answer. The method normally yields a single answer when several might be preferable or needed.

Recently, a number of approaches have been taken based on computing regressions for all possible subsets. Some of these techniques depend heavily on the use of the statistic:

$$C(p) = \frac{RSS(p)}{\hat{\sigma}^2} + 2p - n,$$

where

n = number of measurements,

p = number of variables in the subset,

RSS(p) = residual sum of squares using p variables,

$\hat{\sigma}^2$  = an independent estimate of the error variance.

Usually, an independent estimate of  $\sigma^2$  is not available, and the residual variance from the full model with (k) variables is used for  $\hat{\sigma}^2$ .

If the model is satisfactory, C(p) will be approximately equal to p, subject to sampling variability.

This section describes the BESTCP instruction which computes the ten smallest C(p) statistics among all possible C(p) statistics with k variables and a constant term in the full model.

```
:  
: BESTCP of y in col (C) weights (E) with (k) variables in columns (C)...(C) :  
:
```

The structure of a BESTCP instruction is similar to that of a FIT instruction without storage. The instruction uses a very efficient algorithm to compute only as many regressions as are needed in order to produce the ten smallest C(p) statistics. The results may be thought of as being printed in four sections.

First, a two-line title gives the column number (or column heading) for the independent variable (1st argument), the number of predictor variables (3rd argument) and the number of measurements with non-zero weights. The (k) variables are numbered consecutively from 1 to k and immediately below the title, a table is given relating the variables to the column numbers (or column headings).

Second, under titles REGRESSION WITH 1 VARIABLE, REGRESSIONS WITH 2 VARIABLES, etc., the ten smallest C(p) statistics are printed. There are three columns under each of these C(p) statistics titled VARIABLE, COEFFICIENT, and F RATIO. The first column gives the specific variables in the regression and the second column gives the least squares estimates of the regression coefficients. The F-ratios have 1 and (n-p-1) degrees of freedom. The value of the F-ratio is influenced by the order of the variables in the regression.

Third, all additional C(p) statistics, obtained as a byproduct of computing the ten smallest, are printed with titles C(P) STATISTIC and VARIABLES for REGRESSION WITH 1 VARIABLE, REGRESSIONS WITH 2 VARIABLES, etc.

Fourth, at the bottom, the number of regressions needed to find the ten smallest C(p) statistics is printed along with the number of operations.

### Use of Results

C(p) statistics which are approximately equal to p should be looked for. Remember, p equals the number of variables plus one for the constant term. Among subsets with comparable C(p) statistics, the subsets with the smallest number of variables are usually the most desirable. In a particular problem, there may be practical considerations that should force a variable to be included in the chosen subset. A good analyst will look for patterns among the low and high C(p) statistics. In the example below it should be noticed that variable 1 does not appear in any of the subsets with C(p) statistics greater than 7.

The subset with the smallest C(p) statistic should not necessarily be selected. The small value may be the result of sampling variability. A plot of C(p) versus p can be very helpful in subset selection.

### Example

A set of instructions is listed below to produce C(p) statistics for Hald's data discussed on pages 89 and 90 of Daniel and Wood (1971). The results are shown on pages 142 and 143.

```
WIDTH 80
READ DATA INTO COLUMNS 11 12 13 14 AND 1
  7 26 6 60 78.5
  1 29 15 52 74.3
11 56 8 20 104.3
11 31 8 47 87.6
  7 52 6 33 95.9
11 55 9 22 109.2
  3 71 17 6 102.7
  1 31 22 44 72.5
  2 54 18 22 95.1
21 47 4 26 115.9
  1 40 23 34 83.8
11 66 9 12 113.3
10 68 8 12 109.4
HEAD COL 14/ PCT. C2S
BESTCP 1 1.0 4 11 12 13 14
```

The coefficients printed below C(P) = 5.496 are the 2nd and 3rd coefficients that would be obtained by using the instruction

```
FIT 1, 1.0, 3, 10, 11, 14
```

with 1.0 in every row of column 10. The F ratios are the same as the F-ratios printed by FIT in the third column from the right in the analysis of variance given on the third page.

Remarks

(a) If  $k$  is very large, it takes a lot of computing time to compute all possible regressions. The algorithm used by a BESTCP instruction is believed to be the most efficient one available. If the number of variables is ten or less, the computing time should not be very important. However, if the number of variables is large the computing time could run into minutes depending upon the number of variables, the data and the computer. The number of operations at the end of the printing may be helpful in assessing time requirements.

(b) The BESTCP instruction is an exception to the rule in OMNITAB development that accuracy comes before speed. The results should not be expected to be as accurate as results obtained from an excellent program designed to perform calculations for a single regression. We believe the accuracy of the results is more than adequate for their purpose and have no evidence of numerical instability. As a precaution, a FIT instruction could be used for the subsets under final consideration. This should be done anyway, because further analysis, including an examination of the residuals, should be done before the final selection is made.

(c) In the above example, all possible  $C(p)$  statistics were printed (excluding the one for the regression with no variables). This will not always be the case; particularly if  $k$  is large. If  $m$   $C(p)$  statistics are printed in addition to the ten smallest, these  $m$  statistics will not necessarily be the  $m$  smallest of the remaining  $C(p)$  statistics. They are merely the  $\overline{C}(p)$  statistics which were obtained as a by-product in computing the regressions necessary to obtain the ten smallest.

(d) It is believed that the BESTCP instruction as it now stands is very powerful and useful. It represents the best state of the art computing for regression subset selection. However, with more experience and time it can and should be improved.

Restrictions

(a) The  $C(p)$  statistics are only computed for a regression model with a constant term, but a column number for the constant term must not be in the instruction.

(b) The number of variables (3rd argument of instruction) must be greater than or equal to 3 and less than or equal to 28. If the number of variables is 3, the instruction only finds the seven smallest  $C(p)$  statistics.

(c) Weights can only be equal to 0.0 or 1.0. If the second argument is a column number, all the numbers in the column should be equal to 0.0 or 1.0.

(d) For a given number of variables in a regression (subset size) a maximum of ten  $C(p)$  statistics are printed.

C(P) STATISTICS FOR COLUMN 1 AS A LINEAR FUNCTION OF A CONSTANT AND FROM ONE UP TO 4 VARIABLES WITH 13 MEASUREMENTS WITH NONZERO WEIGHTS

VARIABLE 1 IS COLUMN 11 , VARIABLE 2 IS COLUMN 12 ,  
 VARIABLE 3 IS COLUMN 13 , VARIABLE 4 IS PCT. C2S

REGRESSION WITH 1 VARIABLE

C(P) STATISTIC	VARIABLES
138.731	4
142.486	2
202.549	1
315.154	3

REGRESSIONS WITH 2 VARIABLES

C(P) = 2.678		
VARIABLE	COEFFICIENT	F RATIO
1	1.4683058	146.523
2	.66225047	208.582

C(P) = 5.496		
VARIABLE	COEFFICIENT	F RATIO
1	1.4399583	108.224
4	-.61395362	159.295

C(P) = 22.373		
VARIABLE	COEFFICIENT	F RATIO
3	-1.1998512	40.295
4	-.72460013	100.357

C(P) = 62.438		
VARIABLE	COEFFICIENT	F RATIO
2	.73132958	36.683
3	-1.0083862	11.816

C(P) = 138.226		
VARIABLE	COEFFICIENT	F RATIO
2	.31090473	.172
4	-.45694190	.431

C(P) STATISTIC	VARIABLES
198.094	1 3

REGRESSIONS WITH 3 VARIABLES

C(P) = 3.018		
VARIABLE	COEFFICIENT	F RATIO
1	1.4519380	154.008
2	.41610947	5.026
4	-.23654049	1.863

C(P) = 3.041		
VARIABLE	COEFFICIENT	F RATIO
1	1.6958901	68.716
2	.65691487	220.547
3	.25001752	1.832

C(P) = 3.497

VARIABLE	COEFFICIENT	F RATIO
1	1.0518541	22.113
3	-.41004334	4.236
4	-.64279614	208.240

C(P) = 7.337

VARIABLE	COEFFICIENT	F RATIO
2	-.92341541	12.427
3	-1.4479710	96.940
4	-1.5570444	41.654

REGRESSIONS WITH 4 VARIABLES

C(P) = 5.000

VARIABLE	COEFFICIENT	F RATIO
1	1.5510969	4.337
2	.51016177	.497
3	.10190346	.018
4	-.14406672	.041

14 REGRESSIONS                      56 OPERATIONS

#### 4.6+ Correlation.

We failed to indicate the maximum size for CORRELATION in the "OMNITAB II User's Reference Manual." For the number of variables,  $p$ , the maximum number of measurements,  $n$  (NRMAX), allowed is given in the table below. Note, that the maximum value of  $p$  is 51 and that the worksheet would have to be redimensioned to accommodate the maximum number of measurements. The table assumes the standard worksheet size (and scratch area, 13,500).

$p$	$n$	$p$	$n$	$p$	$n$	$p$	$n$	$p$	$n$
		11	1106	21	589	31	388	41	279
2	3372	12	1019	22	561	32	375	42	270
3	3347	13	945	23	536	33	362	43	262
4	2676	14	880	24	512	34	349	44	254
5	2229	15	823	25	491	35	338	45	247
6	1909	16	773	26	471	36	327	46	240
7	1668	17	728	27	452	37	316	47	233
8	1481	18	688	28	435	38	306	48	226
9	1331	19	651	29	418	39	296	49	219
10	1209	20	619	30	403	40	287	50	213
								51	207

If the number of measurements exceeds the number in the above table for a particular value of  $p$ , the following fatal error message is printed:

INSUFFICIENT SCRATCH AREA.

#### 4.7.1 Probability Plotting.

CAUCHY PLOT, EXTREME PLOT, LAMBDA PLOT, NORMAL PLOT, UNIFORM PLOT,	DEXPONENTIAL PLOT, GAMMA PLOT, LOGISTIC PLOT, PARETO PLOT, WEIBULL PLOT	EXPONENTIAL PLOT, HALFNORMAL PLOT, LOGNORMAL PLOT, POISSON PLOT,
--	---	---

#### Background.

A probability plot is a graphical tool for assessing the goodness of fit of some hypothesized statistical distribution (e.g., normal, Poisson, uniform, etc.) to some observed data. It is a plot of the ordered observations,  $X_i$ , vertically, versus the theoretical "typical values",  $M_i$ , for the ordered observations of the hypothesized distribution, horizontally. The  $X_i$  values plotted along the vertical axis are only a function of the observed data. The  $M_i$  values plotted along the horizontal axis are only a function of the number of observations,  $n$ , and the hypothesized distribution. The discriminatory power of a probability plot stems from the fact that the horizontal axis values,  $M_i$ , are strongly affected by the hypothesized distribution. If the user's choice of a distribution is correct, the  $X_i$  and  $M_i$  will be (approximately) linearly related and the resulting probability plot will be close to a straight line. However, if the user's choice of a distribution is not in accord with the true underlying distribution of the data, then the  $X_i$  and  $M_i$  will not have a simple linear relation and the resulting probability plot will be non-linear.

#### Advantages.

A probability plot is a simple, yet powerful technique for assessing the form of the underlying distribution of an observed set of data. An analysis of a probability plot will help answer the following questions.

- (1) Are the data normal (or exponential, or ...)?

- (2) Are the data longer-tailed or shorter-tailed than normal?
- (3) Are certain observations outliers?
- (4) Is the experiment under "statistical control?"
- (5) Is the assumed regression model correct (an incorrect model will frequently result in gross non-normality of the residuals)?
- (6) Are the  $t$  and  $F$  tests commonly employed in regression and ANOVA valid (they are fully valid only if the underlying distribution of residuals is normal)?
- (7) What are the optimal (minimum variance) estimators for location and dispersion (the optimal estimators are dependent on the actual underlying distribution)?

#### Comparison with other Techniques.

There are several advantages of using a probability plot as opposed to other techniques for checking distributional assumptions.

- (1) It is easy to use and interpret.
- (2) It is a graphical technique and thus it makes use of and presents information about every element in the data set (outliers, for example, stand out). In this respect, probability plots are superior to methods which reduce all of the data to a single statistic (like chi-squared).
- (3) It needs no a priori estimate of the unknown location and dispersion parameters of the distribution (as does, e.g., chi-squared).
- (4) It uses individual data points and therefore avoids class width and number-of-classes problems intrinsic to, e.g., histogram methods.
- (5) It provides "free" estimates of location and dispersion parameters via the intercept and slope of the probability plot of the fitted line to the probability plot.
- (6) It is generally applicable to a wide gamut of distributions, distributional families, and distributional types (continuous and discrete).
- (7) When the hypothesized distribution is inappropriate, the plot provides "feedback" information for a better hypothesized distribution.

#### Interpreting Probability Plots.

The dominant feature to look for in a probability plot is linearity. If the plot approximates a straight line (necessarily from the lower left corner to the upper right corner), then the hypothesized distribution yields a good fit to the observed data. On the other hand, the more non-linear the probability plot, the worse the distributional fit.

As alluded to previously, under probability plot advantages, an important property of a probability plot is the existence of feedback information. If the choice of a distribution results in a non-linear plot, then the type of non-linearity can be used to determine a better-fitting choice of a distribution. To be specific, suppose the choice was a particular symmetric distribution with "moderate" tails (like, say, the normal distribution) and the resulting probability plot was non-linear. If the shape of the plot is vertical, then horizontal, then vertical (VHV), then this suggests that the distribution is not long-tailed enough, and a distribution which is longer-tailed will probably give a better fit. Conversely, if we had a horizontal, then vertical, then horizontal (HVH) type probability plot, then this suggests that a distribution which is shorter-tailed will fit the data

better. Finally, if a probability plot has a non-symmetric cross-over (inflection) point (or no cross-over (inflection) point at all), then this suggests that a non-symmetric (skewed) distribution would be an improved choice for fitting the data.

If the hypothesized distribution adequately describes the data, then the probability plot may show considerable variation about the straight line when the number of measurements is small. Such variation will, of course, tend to decrease as the number of measurements increases.

#### Estimates.

The probability plot correlation coefficient is a measure of the linearity of the probability plot and hence the adequacy of the distributional fit. See FILLIBEN (1975) for percentage points of the PPCC for the normal case.

The intercept and slope of the probability plot serve as estimators of the location parameter  $\mu$  and the scale parameter  $\sigma$  of the distribution, respectively. It is to be noted that in general  $\mu$  and  $\sigma$  do not always correspond to the population mean and standard deviation of the distribution; however, in the important normal case, such equivalence does exist.

A statistician should be consulted for help in interpreting the probability plot correlation coefficient, intercept or slope.

#### Use of Probability Plots.

It is appropriate to apply probability plots not only to "raw" data, but also to residuals from a regression or an analysis of variance. In using OMNITAB probability plots, the user need not order the observations (or residuals) in the column of interest. Such ordering is done automatically and internally. After a probability plot has been formed, the column being operated on remains unchanged. The column is exactly as it was before the probability plot instruction was executed.

The complete form of all the probability plot instructions is given below.

```
:-----:
: CAUCHY PLOT of column (C)      :
:-----:
:                               :
```

```
:-----:
: DEXPONENTIAL PLOT of column (C) :
:-----:
:                               :
```

```
:-----:
: EXPONENTIAL PLOT of column (C)  :
:-----:
:                               :
```

```
:-----:
: EXTREME PLOT of column (C)      :
:-----:
:                               :
```

```
-----
/ EXTREME PLOT with parameter (K) of column (C) \
-----
```

```

: GAMMA PLOT with parameter (K) of column (C) :
:
:
:
: HALFNORMAL PLOT of column (C) :
:
:
:
: LAMBDA PLOT with parameter (K) of column (C) :
:
:
:
: LOGISTIC PLOT of column (C) :
:
:
:
: LOGNORMAL PLOT of column (C) :
:
:
:
: NORMAL PLOT of column (C) :
:
:
:
: PARETO PLOT with parameter (K) of column (C) :
:
:
:
: POISSON PLOT with parameter (K) of column (C) :
:
:
:
: UNIFORM PLOT of column (C) :
:
:
:
: WEIBULL PLOT with parameter (K) of column (C) :
:
:

```

The nine instructions with first word CAUCHY, DEXPONENTIAL, EXPONENTIAL, EXTREME, HALFNORMAL, LOGISTIC, LOGNORMAL, NORMAL and UNIFORM have a single argument; the column number where the data are stored. The five instructions with first word GAMMA, LAMBDA, PARETO, POISSON and WEIBULL have an additional argument, which appears before the column number, to specify the value of the parameter of the distribution. The optional form of EXTREME PLOT also has two arguments for specifying the parameter value and the column number. The main form of EXTREME PLOT gives an extreme value type 1 probability plot. The optional form of EXTREME PLOT, with a parameter specified, gives an extreme value type 2 probability plot. The parameter of a LAMBDA PLOT instruction can be negative, zero, or

positive. The parameter for the other instructions must always be positive. If the parameter value is zero or negative, the following fatal error message is printed:

THE PROBABILITY DISTRIBUTION PARAMETER IS NOT ALLOWED.

Three different sizes of probability plots are possible. The vertical axis always has the same length and the usual number of 51 plotting positions. (We expect to provide a new instruction for varying this size in the near future.) The size of the horizontal axis is determined by the use or non-use of a WIDTH instruction. (See Section C1.1.1.) The three possibilities are:

<u>Case</u>	<u>(n) of WIDTH (n)</u>	<u>Number of Plotting Positions</u>
1	120	51 x 101
2	72-119	51 x 61
3	59-71	51 x 41

If a WIDTH instruction has not been used,  $n = 120$  and the plot is the normal size, i.e.,  $51 \times 101$ . If (n) of WIDTH (n) is less than 59, a plot is not given and the following informative diagnostic is printed:

THE INSTRUCTION WAS IGNORED BECAUSE ...  
WIDTH = (n) IS TOO SMALL OR TOO LARGE.

The one line title at the top of the plot and the one line of information at the bottom of the plot are somewhat different for each of the three different sized plots. For a Weibull probability plot with parameter 2.0 and 250 measurements in column 45, the three possible headings are:

WEIBULL PROBABILITY PLOT OF 250 MEASUREMENTS IN COLUMN 45 WITH PARAMETER = 2.0000000

WEIBULL PR. PLOT OF COLUMN 45 N = 250 PARAMETER 2.0000000

WEIBULL PR PLOT OF COLUMN 45 PARAM. 2.0000000

The line of information at the bottom of the plot, in each case, is:

PROB. PLOT CORR. COEFF. = .9952, ESTIMATES \* INTERCEPT = -1.95270000, SLOPE 2.2013169

PLOT COR COEF = .9952, EST\* INT. = 1.9527000, SLOPE = 2.2013169

PROBABILITY PLOT CORRELATION COEFF. = .9952

Note, if the WIDTH is less than 72, the number of measurements is not given in the heading and estimates of the intercept and slope are not printed at the bottom.

#### Caution.

In addition to the two-word commands EXPONENTIAL PLOT and EXTREME PLOT, we have had for some time the one-word commands EXPONENTIAL and EXTREMA. When these one-word commands are used PLOT should not be a noise word immediately following the command or the instruction will be misinterpreted as a probability plot instruction. (PLOT can be a noise word in the instruction just as long as it does not appear immediately after the command without an intervening argument or noise word.)

Example.

The following set of instructions is used to produce the normal probability plot of 200 random normal deviates which appears on the next page. A uniform probability plot on page 151 and a Cauchy plot on page 152 are also given for comparison. Since the data are random normal deviates, the normal probability plot is approximately a straight line. The uniform distribution has shorter tails, and hence the probability plot does not resemble a straight line and has a vertical-horizontal-vertical appearance. The tails of a Cauchy distribution are longer than the tails of a normal distribution and hence the probability plot has a horizontal-vertical-horizontal appearance.

OMNITAB 10/24/75 EXAMPLE OF NORMAL PROBABILITY PLOT

LIST OF COMMANDS, DATA AND DIAGNOSTICS

SET TEST DATA IN COLUMN 1

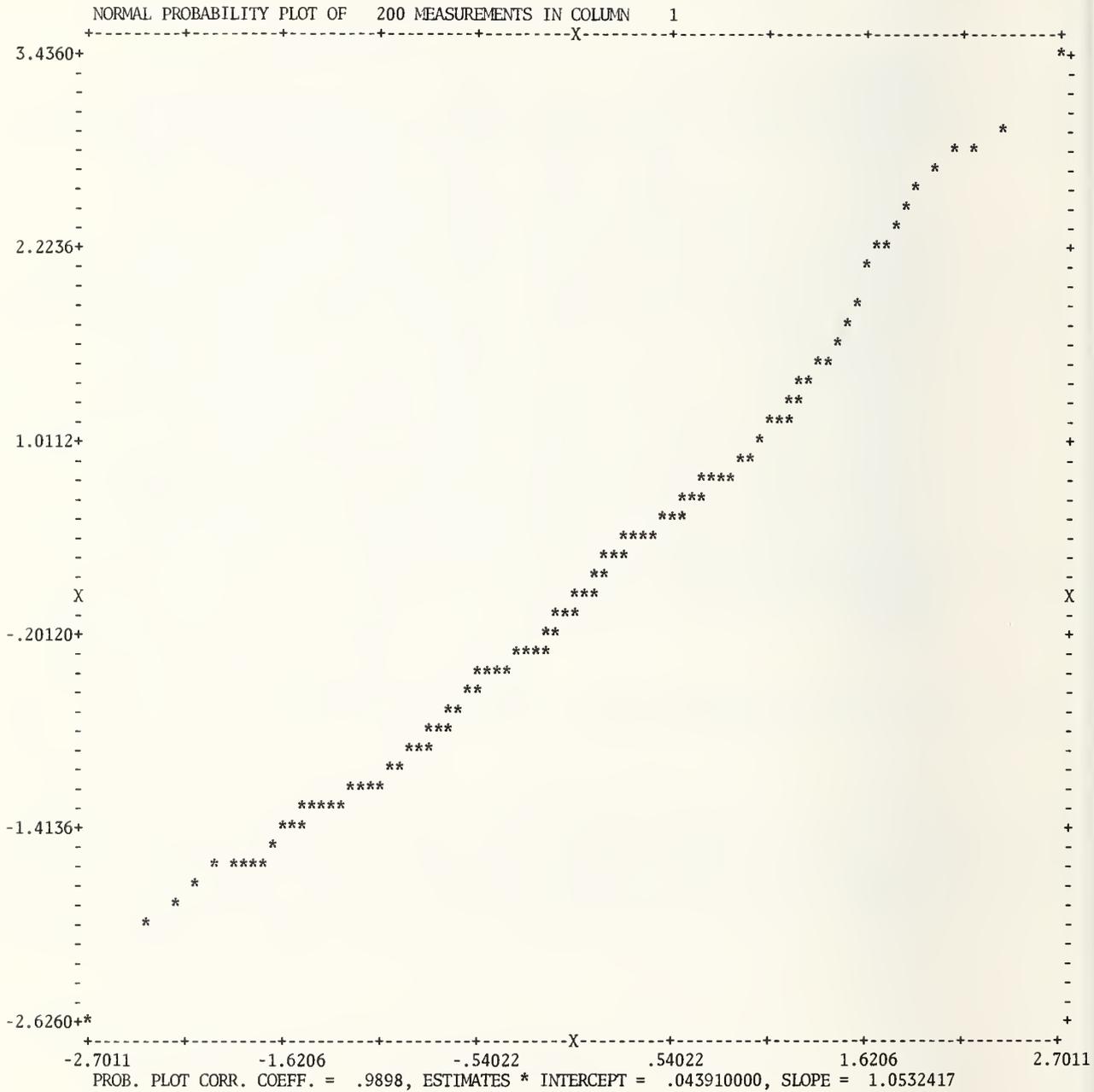
-1.276	-1.218	-0.453	-0.350	0.723	0.676	-1.099	-0.314	-0.394	-0.633
-0.318	-0.799	-1.664	1.391	0.382	0.733	0.653	0.219	-0.681	1.129
-1.377	-1.257	0.495	-0.139	-0.854	0.428	-1.322	-0.315	-0.732	-1.348
2.334	-0.337	-1.955	-0.636	-1.318	-0.433	0.545	0.428	-0.297	0.276
-1.136	0.642	3.436	-1.667	0.847	-1.173	-0.355	0.035	0.359	0.930
0.414	-0.011	0.666	-1.132	-0.410	-1.077	0.734	1.484	-0.340	0.789
-0.494	0.364	-1.237	-0.044	-0.111	-0.210	0.931	0.616	-0.377	-0.433
1.048	0.037	0.759	0.609	-2.043	-0.290	0.404	-0.543	0.486	0.869
0.347	2.816	-0.464	-0.632	-1.614	0.372	-0.074	-0.916	1.314	-0.038
0.637	0.563	-0.107	0.131	-1.808	-1.126	0.379	0.610	-0.364	-2.626
2.176	0.393	-0.924	1.911	-1.040	-1.168	0.485	0.076	-0.769	1.607
-1.185	-0.944	-1.604	0.185	-0.258	-0.300	-0.591	-0.545	0.018	-0.485
0.972	1.710	2.682	2.813	-1.531	-0.490	2.071	1.444	-1.092	0.478
1.210	0.294	-0.248	0.719	1.103	1.090	0.212	-1.185	-0.338	-1.134
2.647	0.777	0.450	2.247	1.151	-1.676	0.384	1.133	1.393	0.814
0.398	0.318	-0.928	2.416	-0.936	1.036	0.024	-0.560	0.203	-0.871
0.846	-0.699	-0.368	0.344	-0.926	-0.797	-1.404	-1.472	-0.118	1.456
0.654	-0.955	2.907	1.688	0.752	-0.434	0.746	0.149	-0.170	-0.479
0.522	0.231	-0.619	-0.265	0.419	0.558	-0.549	0.192	-0.334	1.373
-1.288	-0.539	-0.824	0.244	-1.070	0.010	0.482	-0.469	-0.090	1.171

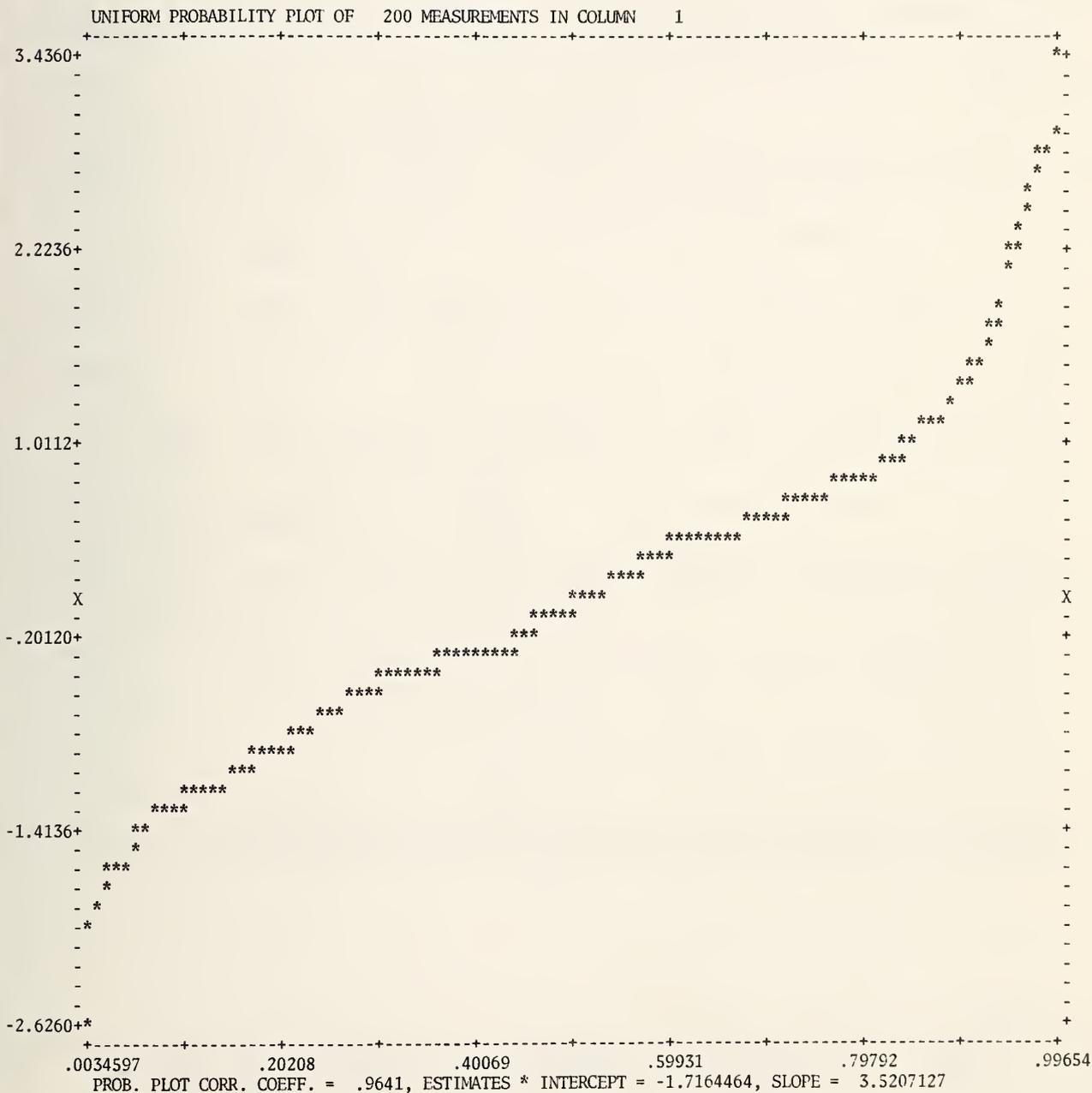
NORMAL PLOT OF RANDOM NORMAL DEVIATES IN COLUMN 1

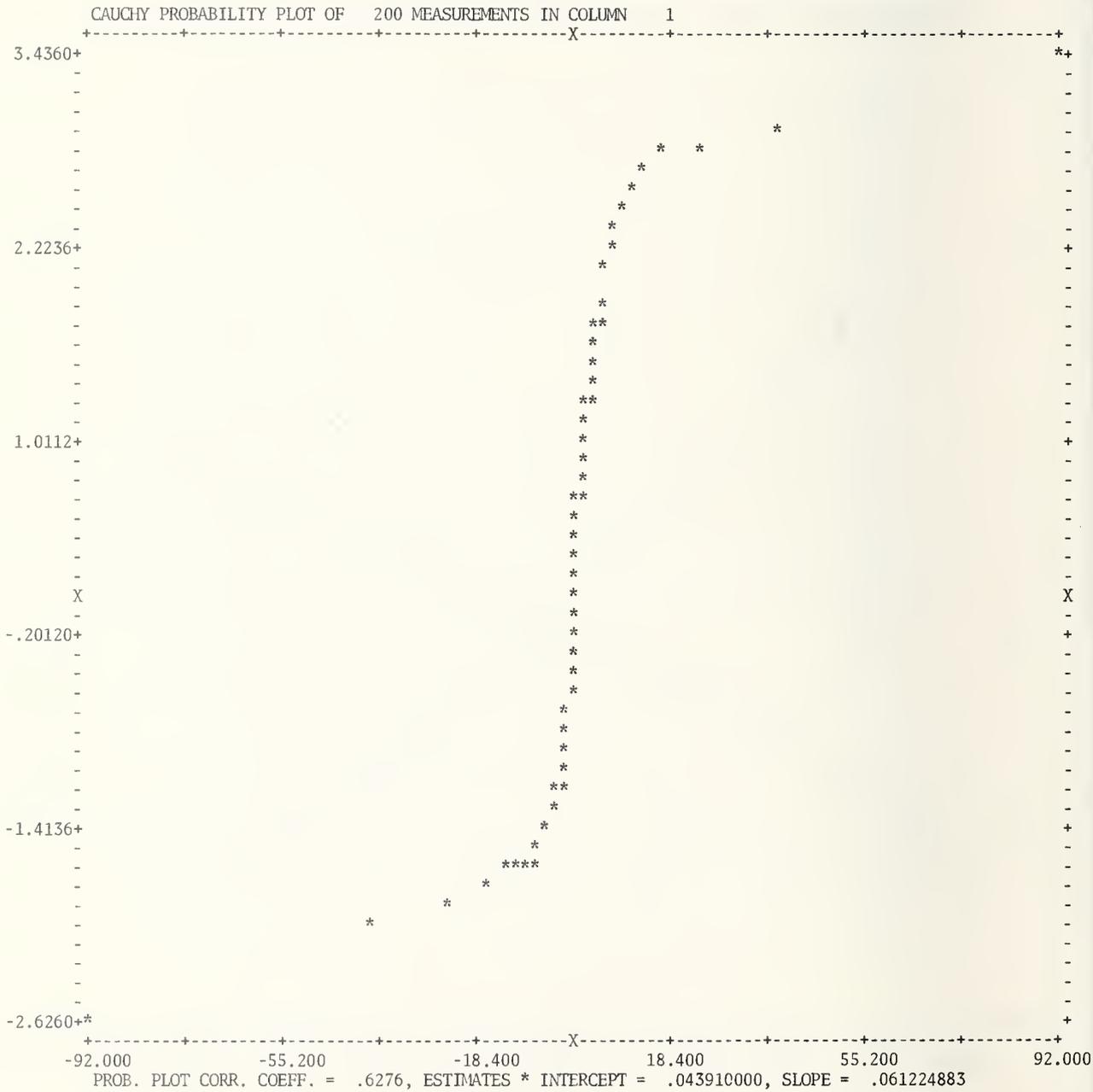
UNIFORM PLOT OF COLUMN 1

CAUCHY PLOT OF COLUMN 1

STOP







#### 4.8+ References.

- DANIEL, CUTHBERT (1959). Use of half-normal plots in interpreting factorial two-level experiments. *Technometrics*, 1, No. 4, 311-341.
- DANIEL, CUTHBERT and WOOD, FRED S. (1971). Fitting Equations to Data. Wiley Interscience.
- FILLIBEN, JAMES J. (1975). The probability plot correlation coefficient test for normality. *Technometrics*, 17, No. 1, 111-117.
- FURNIVAL, GEORGE M. and WILSON, ROBERT W., Jr. (1974). Regressions by leaps and bounds. *Technometrics*, 16, 499-511.
- GARDNER, D.A. and HULL, N.C. (1966). An approximation to two-sided tolerance limits for normal populations. *Technometrics*, 8, 115-122.
- GORMAN, J.W. and TOMAN, R.J. (1966). Selection of variables for fitting equations to data. *Technometrics*, 8, 27-51.
- HAHN, G.J. and SHAPIRO, S.S. (1967). Statistical Methods in Engineering. John Wiley & Sons, Inc. Pages 260-308.
- HOCKING, R.R. (1976). The analysis and selection of variables in linear regression. *Biometrics*, (Invited paper), 32, 1-49.
- JENKINS, G.M. and WATTS, D.G. (1968). Spectral Analysis and its Applications. Holden-Day.
- MCGILL, ROBERT, TUKEY, JOHN W. and LARSON, W.A. (1977). Variations of box plots. Submitted to *American Statistician*.
- NELSON, WAYNE and THOMPSON, V.C. (1971). Weibull probability papers. *Journal of Quality Technology*, 3, No. 2, 45-50.
- NEMENYI, PETER; DIXON, SYLVIA K. and WHITE, NATHANIEL B., JR. (1975). Statistics from Scratch. To be published by Holden-Day.
- SNEDECOR, G.W. (1956). Statistical Methods, 5th Edition, Iowa State College Press.
- TUKEY, J.W. (1970). Exploratory Data Analysis. 3 vols. Preliminary edition. Addison-Wesley.
- TUKEY, J.W. (1972). Statistical Papers in Honor of George W. Snedecor. Edited by T.A. Bancroft. pp 295-296.
- WILK, M.B. and GNANADESIKAN, R. (1968). Probability plotting methods for the analysis of data. *Biometrika*, 55, 1-17.

## 5. NUMERICAL ANALYSIS.

### 5.1\* Special Integrals.

COSINTEGRAL, EEXPINTEGRAL, EINTEGRAL, EXPINTEGRAL,  
HCOSINTEGRAL, HSININTEGRAL, NEGEINTEGRAL, SININTEGRAL

The special integrals described below are frequently encountered in physical problems.

```
-----  
: COSINTEGRAL of (E) put in column (C) :  
-----
```

Computes values of the cosine integral

$$Ci(x) = \gamma + \ln(x) + \int_0^x [(\cos(t)-1)/t]dt, x > 0,$$

where  $\gamma$  = Euler's constant = 0.57721566...

For the numbers 1.0, 2.0, 3.0, and 20.0 in column 1, the instruction

COSINTEGRAL of column 1 put in column 2

would put the numbers .33740392, .42298083, .11962979, and .044419821 into column 2.

If the value(s) specified by (E) is (are) negative or zero, the following arithmetic fault message is printed:

FUNCTION NOT DEFINED FOR NEGATIVE OR ZERO VALUES.

However, negative values, which produce complex answers, are allowed in the optional form with three arguments described below.

```
-----  
/ COSINTEGRAL of (E) put real values in col (C) and imaginary values in col (C) \  
-----
```

Computes values of the cosine integral when x is negative. The result is complex with real and imaginary parts given by

$$Ci(-x) = Ci(x) - \pi i$$

For the numbers -1.0, -2.0, -5.0 and -20.0 in column 3, the instruction

COSINTEGRAL of col 3 put real part in col 7 and imaginary part in col 8

would put the numbers .33740392, .42298083, -.19002975, and .044419821 into column 7 and the numbers -3.1415927, -3.1415927, -3.1415927 and -3.1415927 into column 8.

If any value specified by (E) is zero, the following arithmetic fault message is printed:

FUNCTION NOT DEFINED FOR ZERO VALUES.

:  
: EEXPINTEGRAL for (n) of (E) put in column (C)  
:

Computes values of

$$e^x E_n(x), x \geq 0, n = 0, 1, 2, \dots,$$

where  $E_n(x)$  is described below under EXPINTEGRAL; except n must be greater than 1 when x equals zero.

For the numbers 1.0, 2.0, 3.0, and 20.0 in column 1, the instruction

EEXPINTEGRAL for n = 2 of column 1 put in column 2

would put the numbers .40365263, .27734276, .21374878, and .045629090 into column 2.

If the value(s) specified by (E) is (are) negative, the following arithmetic fault message is printed:

FUNCTION NOT DEFINED FOR NEGATIVE VALUES.

If  $x = 0$  and  $n = 0$  or 1, the following arithmetic fault message is printed:

FUNCTION NOT DEFINED FOR ZERO VALUES.

If (n) is a negative integer, the following fatal error message is printed:

INCORRECT ARGUMENT IN INSTRUCTION.

If (n) is not an integer, the following fatal error message is given:

IMPROPER TYPE OF ARGUMENT.

:  
: EINTEGRAL of (E) put in column (C)  
:

Computes values of the integral

$$Ei(x) = - \int_{-x}^{\infty} [e^{-t}/t]dt, x > 0.$$

For the numbers 1.0, 2.0, 3.0, and 20.0 in column 1, the instruction

EINTEGRAL of column 1 put in column 2

would put the numbers 1.8951178, 4.9542344, 9.9338325, 25,615,653. into column 2.

If the value(s) specified by (E) is (are) negative or zero, the following arithmetic fault message is printed:

FUNCTION NOT DEFINED FOR NEGATIVE OR ZERO VALUES.

:  
: EXPINTEGRAL for (n) of (E) put in column (C)  
:

Computes values of the exponential integral

$$E_n(x) = \int_1^{\infty} t^{-n} e^{-xt} dt; n = 0, 1, 2, \dots, x \geq 0;$$

except n must be greater than 1 when x equals zero.

For the numbers 1.0, 2.0, 3.0 and 20.0 in column 1, the instruction

EXPINTEGRAL for n = 2 of column 1 put in column 2

would put the numbers .14849551, .037534262, .010641925, and 9.4048564-11 into column 2.

If the value(s) specified by (E) is (are) negative, the following arithmetic fault message is printed:

FUNCTION NOT DEFINED FOR NEGATIVE VALUES.

If x = 0 and n = 0 or 1, the following arithmetic fault is printed:

FUNCTION NOT DEFINED FOR ZERO VALUES.

If (n) is a negative integer, the following fatal error message is printed:

INCORRECT ARGUMENT IN INSTRUCTION.

If (n) is not an integer, the following fatal error message is printed:

IMPROPER TYPE OF ARGUMENT.

:  
: H COSINTEGRAL of (E) put in column (C)  
:

Computes values of the hyperbolic cosine integral

$$\text{Chi}(x) = \gamma + \ln(x) + \int_0^x [(\cosh(t)-1)/t] dt, x > 0,$$

where  $\gamma$  = Euler's constant = 0.57721566...

For the numbers 1.0, 2.0, 3.0, and 20.0 in column 1, the instruction

HCOSINTEGRAL of column 1 put in column 2

would put the numbers .83786694, 2.4526669, 4.9603921, and 12,807,826. into column 2.

If the value(s) specified by (E) is (are) negative or zero, the following arithmetic fault message is printed:

FUNCTION NOT DEFINED FOR NEGATIVE OR ZERO VALUES.

However, negative values, which produce complex answers, are allowed in the optional form with three arguments described below.

HCOSINTEGRAL of (E) put real part in col (C) and imaginary part in col (C)

Computes values of the hyperbolic cosine integral when x is negative. The result is complex with real and imaginary parts given by

$$\text{Chi}(-x) = \text{Chi}(x) - \pi i.$$

For the numbers -1.0, -2.0, -5.0 and -20.0 in column 3, the instruction

HCOSINTEGRAL of col 3 put real part in col 7 and imaginary part in col 8

would put the numbers .83786694, 2.4526669, 20.092063, and 12,807,826. into column 7 and the numbers -3.1415927, -3.1415927, -3.1415927 and -3.1415927 into column 8.

If any x-value is zero, the following arithmetic fault message is printed:

FUNCTION NOT DEFINED FOR ZERO VALUES.

HSININTEGRAL of (E) put in column (C)

Computes values of the hyperbolic cosine integral

$$\text{Shi}(x) = \int_0^x [\sinh(t)/t]dt.$$

When x is negative, the integral is evaluated as  $\text{Shi}(-x) = -\text{Shi}(x)$ .

For the numbers 1.0, -1.0, 2.0, 3.0, and 20.0 in column 1, the instruction

HSININTEGRAL of column 1 put in column 2

would put the numbers 1.0572509, -1.0572509, 2.5015674, 4.9734405, and 12,807,826. into column 2.

NEGEINTEGRAL of (E) put in column (C)

Computes values of the integral

$$e^{-x}\text{Ei}(x), x > 0,$$

where  $\text{Ei}(x)$  is described above under EINTEGRAL.

For the numbers 1.0, 2.0, 3.0, and 20.0 in column 1, the instruction

NEGEINTEGRAL of column 1 put in column 2

would put the numbers .69717488, .67048271, .49457640, and .052797795 into column 2.

If the value(s) specified by (E) is (are) not positive, the following arithmetic fault message is printed:

FUNCTION NOT DEFINED FOR NEGATIVE OR ZERO VALUES.

```
:  
: SININTEGRAL of (E) put in column (C)  
:
```

Computes values of the sine integral

$$\text{Si}(x) = \int_0^x [\sin(t)/t]dt.$$

When x is negative, the integral is evaluated as  $\text{Si}(-x) = -\text{Si}(x)$ .

For the numbers 1.0, -1.0, 2.0, 3.0, and 20.0 in column 1, the instruction

SININTEGRAL of column 1 put in column 2

would put the numbers .94608307, -.94608307, 1.6054130, 1.8486525, and 1.5482417 into column 2.

#### References.

STEGUN, IRENE A and ZUCKER, RUTH (1974). Automatic computing methods for special functions. Part II. The exponential integral  $E_n(x)$ . J. of Res., B, NBS, 74B, 199-216.

STEGUN, IRENE A and ZUCKER, RUTH (1976). Automatic computing methods for special functions. Part III. The sine, cosine, exponential integrals and related functions. J. of Res., B, NBS, 80B, 291-311.

#### 5.1.1 Differences.

DIFFERENCES,       SDIFFERENCES,  
DIVDIFFERENCES,   SDIVDIFFERENCES

Differences and, in particular, divided differences are used extensively in numerical analysis for studying the functional form of data, interpolation, inverse interpolation, numerical integration and numerical differentiation.

```
:  
: DIFFERENCES of y in column (C)  
:
```

The DIFFERENCES instruction automatically prints a table of the 1st, 2nd, ..., 7th differences; providing NRMAX is at least eight. If NRMAX is less than eight, the instruction prints the maximum number of differences; namely (NRMAX-1). The numbers in the first column are printed on every other line to produce a more readable table.

Ignoring the OMNITAB line at the top of the page, the instructions

```
SET Y IN COLUMN 2  
0, .20134, .30452, .52110  
DIFFERENCES OF COLUMN 2
```

would produce the following printing:

COLUMN Y	1ST	2ND	3RD
0.			
.20134000	.20134000		
.30452000	.10318000	-.098159997	.21156000
.52110000	.21658000	.11340000	

The numbers in the designated column appear under the title Y. The appropriate column heading appears above Y. Any entry in the table on the right of the first column is formed by subtracting the number just above it on the left from the number just below it on the left. For example,  $.11340000 = .21658000 - .10318000$ .

DIFFERENCES of y in column (C) put in columns (C), (C), ... , (C)

Similar to the above form; except differences are stored in the worksheet in consecutive rows starting with row 1. The number of arguments determines the number of differences which are stored. The first differences are stored in the column designated by the 2nd argument, the second differences are stored in the column designated by the 3rd argument, etc. and the kth differences are stored in the column designated by the last argument. The value of k must be less than NRMAX. The appropriate column headings are printed above the titles 1ST, 2ND, etc.

: SDIFFERENCES of y in column (C) put in columns (C), (C), ... , (C) :

Same as the above instruction; except the automatic printing is suppressed. If the number of arguments is one, the following informative diagnostic is printed:

THE INSTRUCTION WAS IGNORED BECAUSE ...  
COMMAND BEGINS WITH S AND STORAGE MUST BE REQUESTED.

: DIVDIFFERENCES for x in col (C) and y in col (C) :

The DIVDIFFERENCES instruction automatically prints a table of the 1st, 2nd, ..., 6th divided differences; providing NRMAX is at least seven. If NRMAX is less than seven, the instruction prints the maximum number of divided differences, namely (NRMAX-1). The numbers in the first two columns for x and y are printed on every other line to produce a more readable table.

Ignoring the OMNITAB line at the top of the page, the instructions

```
SET X IN COLUMN 1
0, .2, .3, .5
SET Y IN COLUMN 2
0 .20134, .30452, .52110
DIVDIFFERENCES FOR X IN COLUMN 1 AND Y IN COLUMN 2
```

would produce the following printing:

COLUMN X	1	COLUMN Y	2	1ST	2ND	3RD
0.		0.		1.0067000		
.20000000		.20134000		1.0318000	.083666841	
.30000000		.30452000		1.0829000	.17033313	.17333259
.50000000		.52110000				

The first two columns of the table give the values of x and y in the columns designated by the first two arguments of the instruction. The jth entry in the column for the ith order divided differences is computed as:

$$f_{i,j} = (f_{i-1,j+1} - f_{i-1,j}) / (x_{i+j} - x_j), \text{ with } f_{0,j} = y_j.$$

For example, with i=3 and j=1,

$$\begin{aligned} f_{3,1} &= (f_{2,2} - f_{2,1}) / (x_4 - x_1), \\ &= (.17033313 - .083666841) / (0.5 - 0.0), \\ &= .17333259 \end{aligned}$$

DIVDIFFERENCES for x in col (C) and y in col (C) put in cols (C), (C), ... , (C)

Similar to the above form; except divided differences are stored in the worksheet. The number of arguments determines the number of divided differences which are stored. The first divided differences are stored in the column designated by the 3rd argument, the second divided differences are stored in the column designated by the 4th argument, etc. and the kth divided differences are stored in the column designated by the last argument. The value of k must be less than (NRMAX-1). The appropriate column headings are printed above the titles 1ST, 2ND, etc.

SDIVDIFFERENCES for x in (C) and y in (C) put in cols (C), (C), ... , (C)

Same as the above form; except the automatic printing is suppressed. If the number of arguments is two, the following informative diagnostic is printed:

INSTRUCTION WAS IGNORED BECAUSE ...  
COMMAND BEGINS WITH S AND STORAGE MUST BE REQUESTED.

#### References.

- HAMMING, R. W. (1962). Numerical Methods For Scientists And Engineers. McGraw-Hill Book Company, Inc. Pages 3-13 and 98-101; in particular.
- HILDEBRAND, F. B. (1956). Introduction To Numerical Analysis. McGraw-Hill Book Company, Inc. Chapter 2; in particular.

## 7. ARRAY OPERATIONS.

### 7.4\* Properties Of An Array.

Minor changes have been made in the printing of the APROPERTIES instructions. A maximum of 71 characters is printed on any line.

## 8. MATRIX OPERATIONS.

### 8.5\* Matrix Analysis.

New algorithms are now being used for MINVERT, MEIGEN and SOLVE which provide substantially improved accuracy. The MINVERT and SOLVE instructions use routines given in:

FORSYTHE, G.E. and MOLER, C.B. (1967) Computer Solution of Linear Algebraic Systems. Prentice-Hall.

The MEIGEN instruction uses routines given in:

SMITH, B.T., BOYLE, J.M., GARBOW, B.S., IKEBE, Y., KLEMA, V.C. and MOLER, C.B. (1974). Matrix Eigensystem Routines - EISPACK Guide. Lecture Notes in Computer Science, 6, Springer-Verlag.

The value  $r=77$  on page 212 of NBS TN 552 for the largest matrix which can be inverted has been increased to 81 (81 is the solution of  $r^2 + 4r = 13,500$ ). (To obtain square matrices larger than  $62 \times 62$  but less than  $82 \times 82$ , the worksheet would have to be redimensioned.)

The restriction on the size of matrices which can be used with MEIGEN given on page 211 of NBS TN 552 has been removed. MEIGEN can now be used with any symmetric matrix which will fit in the worksheet.

Note carefully, a MEIGEN instruction which stores the eigenvalues in the worksheet, now stores the eigenvalues in ascending order. However, if any eigenvalue is not found, the eigenvalues are not stored in ascending order and the following informative diagnostic message is printed

ITERATION FAILED TO FIND AN EIGENVECTOR. (n) UNORDERED VALUES FOUND.

The entries in the matrix shown at the bottom of page 211 of NBS TN 552 are incorrect and should read:

<u>Row/Column</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>
<u>2</u>	72.0	56.0	-64.0	-64.0
<u>3</u>	56.0	72.0	-64.0	-64.0
<u>4</u>	-64.0	-64.0	72.0	56.0
<u>5</u>	-64.0	-64.0	56.0	72.0

The results of using the instruction

MEIGEN matrix A in row 2, col 3 size 4x4, put values in col 7, vectors in 2,8  
 gives the following results instead of those shown at the top of page 212 of NBS TN 552.

Eigenvalues		Eigenvectors				
Row/Column	<u>7</u>	Row/Column	<u>8</u>	<u>9</u>	<u>10</u>	<u>11</u>
<u>1</u>	-1.1802143-07	<u>2</u>	.50000000	-2.3403118-08	.70710678	.50000000
<u>2</u>	15.999998	<u>3</u>	.49999999	-1.5923433-08	-.70710678	.50000000
<u>3</u>	16.000000	<u>4</u>	.49999995	-.70710680	4.4801595-09	-.49999998
<u>4</u>	2.5599998+02	<u>5</u>	.50000003	.70710674	0.	-.50000001

8.6\* Matrix Properties.

Numerous minor changes have been made in the printing produced by a MPROPERTIES instruction. The following two pages show the results of using the set of instructions on page 218 of NBS TN 552 and in effect replace pages 219 and 220 of NBS TN 552. If a MLABEL instruction has been used, the label is used in the title of the printing.

PROPERTIES OF THE 4X4 MATRIX STARTING AT (1,1)

ROW - FOR PROPERTIES STORED IN COLUMN 10

GENERAL

1	TRACE ( 4 VALUES USED)	101.00000
2	TRACE NUMBER TWO	1255.0000
3	MAXIMUM NUMBER	36.000000
4	MINIMUM NUMBER	4.0000000
5	MAXIMUM NUMBER IN ABSOLUTE VALUE	36.000000
6	MINIMUM NUMBER IN ABSOLUTE VALUE	4.0000000
7	MINIMUM NON-ZERO ABSOLUTE VALUE	4.0000000
8	NUMBER OF POSITIVE NUMBERS	16
9	NUMBER OF ZERO NUMBERS	0
10	NUMBER OF NEGATIVE NUMBERS	0
11	SUM OF TERMS	303.00000
12	AVERAGE	18.937500
13	SUM OF SQUARES	7691.0000
14	SUM OF SQUARES ABOUT MEAN	1952.9375
15	WITHIN ROWS SUM OF SQUARES	1112.7500
16	WITHIN COLS SUM OF SQUARES	1112.7500
17	SUM OF ABSOLUTE VALUES	303.00000
18	AVERAGE OF ABSOLUTE VALUES	18.937500

SPECIFIC

19	DETERMINANT	1023.9998
20	RANK	4
	NORMS	
21	SQUARE ROOT OF SUM OF B(I,J)**2	.0000027
22	N*MAXIMUM(B(I,J))	.000012
23	MAXIMUM VALUE OF ROW SUM	.0000025
24	NORMALITY	YES*(1)
25	SYMMETRY	YES*(1)
26	SKEW SYMMETRY	NO*(0)
27	DIAGONALITY	NO*(0)
28	ORTHOGONALITY: A'A = I	NO*(0)
29	A'A = DIAGONAL MATRIX	NO*(0)
30	TRIANGULAR	NO**(0)
31	STOCHASTIC (ROW AND/OR COL SUMS=1)	NO*** (0)

\* IF ANSWER IS NO: (R,C)= 0. IF ANSWER IS YES: (R,C)= 1, IF EXACT OR (R,,C)= 2, IF TOLERANCE IS MET.

TRIANGULAR

\*\* (R,C)=0, IF ANSWER IS NO. (R,C)=1, IF UPPER PART OF MATRIX=0. (R,C)=2, IF LOWER PART=0. (R,C)=3, IF ALL OFF DIAGONAL TERMS=0.

STOCHASTIC

\*\*\* (R,C)=0, IF MATRIX IS NOT STOCHASTIC. (R,C)=1, IF EACH ROW SUM=1. (R,C)=2, IF EACH COL SUM=1. (R,C)=3, IF EACH ROW AND COL SUM=1.

PROPERTIES OF THE 3X4 MATRIX STARTING AT (1,1)

GENERAL

1	TRACE ( 3 VALUES USED)	65.000000
2	TRACE NUMBER TWO	644.00000
3	MAXIMUM NUMBER	36.000000
4	MINIMUM NUMBER	4.0000000
5	MAXIMUM NUMBER IN ABSOLUTE VALUE	36.000000
6	MINIMUM NUMBER IN ABSOLUTE VALUE	4.0000000
7	MINIMUM NON-ZERO ABSOLUTE VALUE	4.0000000
8	NUMBER OF POSITIVE NUMBERS	12
9	NUMBER OF ZERO NUMBERS	0
10	NUMBER OF NEGATIVE NUMBERS	0
11	SUM OF TERMS	200.00000
12	AVERAGE	16.666667
13	SUM OF SQUARES	4666.0000
14	SUM OF SQUARES ABOUT MEAN	1332.6667
15	WITHIN ROWS SUM OF SQUARES	740.00000
16	WITHIN COLS SUM OF SQUARES	829.33334
17	SUM OF ABSOLUTE VALUES	200.00000
18	AVERAGE OF ABSOLUTE VALUES	16.666667

SPECIFIC

28	ORTHOGONALITY: $A^T A = I$	NO*(0)
29	$A^T A =$ DIAGONAL MATRIX	NO*(0)

\* (R,C)=0, IF MATRIX IS NOT ORTHOGONAL.  
 (R,C)=1 OR 2, IF MATRIX IS ORTHOGONAL ROW WISE.  
 (R,C)=3 OR 4, IF MATRIX IS ORTHOGONAL COLUMN WISE.  
 ( (R,C)=1: IF I=1 OR 3, ORTHOGONALITY IS EXACT;  
 IF I=2 OR 4, RELATIVE WITHIN ERROR BOUND OF .1E-6.)

11.+ INDEX OF COMMANDS DESCRIBED IN PART C SUPPLEMENT

An asterisk immediately following the page number indicates that the corresponding instruction has been modified in some way. Otherwise, the command is new.

ALABEL	23	FIXED	26*	PROPORTIONS	50
APROPERTIES	161*	FOURPLOTS	31	RANGE	51
AVERAGE	49*	FRANCAIS	26	READ UNIT "L"	47
BACKSPACE UNIT "L"	46	FULL	17	READ UNIT "L" "L"	47
BESTCP	139	GAMMA PLOT	147	RECODE	58
BRIEF	17	HALFNORMAL PLOT	147	REMOTE	19
CALCOMP AXIS	39	HISTOGRAM	63*	REPLACE	58
CALCOMP FAST	40	HCOSINTEGRAL	156	RETAIN	59
CALCOMP PAPER	39	HSININTEGRAL	157	REWIND UNIT "L"	47
CALCOMP PLOT	36	INTERACTIVE	17	SDIFFERENCES	159
CALCOMP SIZE	39	ITALIANO	26	SDIVDIFFERENCES	160
CALCOMP SLOW	40	JAPANESE	26	SET UNIT "L"	47
CALCOMP SPEED	40	LABEL	23	SKIP UNIT "L"	47
CALCOMP TAPE "L"	41	LAMBDA PLOT	147	SININTEGRAL	158
CAUCHY PLOT	146	LENGTH	18	SLOVENE	26
CENSOR EQ	52	LOCAL	19	SOLVE	162*
CENSOR GE	52	LOGISTIC PLOT	147	SSTEM LEAF	71
CENSOR GT	52	LOGNORMAL PLOT	147	STATISTICAL	96*
CENSOR LE	52	MEDIAN	50	STATPLOTS	72
CENSOR LT	52	MEIGEN	162*	STDDEV	51
CENSOR NE	52	MINVERT	162*	STEM LEAF	64
CHOOSE	54	MLABEL	24	SVENSKA	26
CODE	55	MPROPERTIES	163*	TABLE AVERAGE	80
CORRELATION	144*	NCPLOT	31	TABLE FREQUENCY	80
COSINTEGRAL	154	NEDERLANDS	26	TABLE MAXIMUM	81
CPLOT	26	NEGEINTEGRAL	157	TABLE MEDIAN	81
CREAD UNIT "L"	46	NORMAL PLOT	147	TABLE MINIMUM	81
CREAD UNIT "L" "L"	46	NORSK	26	TABLE PERCENTAGE	81
CSET UNIT "L"	46	NPLOT	32	TABLE PROPORTION	81
DANSK	26	NTABLE AVERAGE	79	TABLE RANGE	81
DAYS	49	NTABLE FREQUENCY	79	TABLE STDDEV	81
DELETE	56	NTABLE MAXIMUM	79	TABLE SUM	81
DEUTSCH	26	NTABLE MEDIAN	79	TERMINAL	20
DEXPONENTIAL PLOT	146	NTABLE MINIMUM	79	TWOPLOTS	32
DIFFERENCES	158	NTABLE PERCENTAGE	79	TWOWAY	116*
DIVDIFFERENCES	159	NTABLE PROPORTION	79	UNIFORM PLOT	147
EEXPINTEGRAL	155	NTABLE RANGE	79	UNIT	48
EINTEGRAL	155	NTABLE STDDEV	80	VOCABULARY	26
ENDFILE UNIT "L"	47	NTABLE SUM	80	WEIBULL PLOT	147
ENGLISH	26	OMIT	57	WIDTH	20
ESPAÑOL	26	ONEWAY	104*	WRITE UNIT "L"	48
EXPINTEGRAL	156	PARETO PLOT	147	WRITE UNIT "L" "L"	48
EXPONENTIAL PLOT	146	PERCENTAGES	50	YUGOSLAV	26
EXTREME PLOT	146	POISSON PLOT	147		
FIT	125*	PORTUGESE	26		

PART D SUPPLEMENT

LIST OF INSTRUCTIONS

LIST OF INSTRUCTIONS

\*\*\*\*\* A \*\*\*\*\*

ALABEL (label,) (R,) (C,) (r,) (c,), (label,) (R,) (C,) (r,) (c,), ... V,C  
 AVERAGE of numbers in cols (C), (C), ..., (C) put in cols (C), (C), ..., (C) E

\*\*\*\*\* B \*\*\*\*\*

BACKSPACE UNIT "L" (r) logical records 1  
 BESTCP of y in col (C) weights (E) with (k) variables in columns (C)...(C) V,P  
 BRIEF \$ no arguments 0

\*\*\*\*\* C \*\*\*\*\*

CALCOMP AXIS (K) height of y-axis, (K) width of x-axis 2  
 CALCOMP FAST or zip mode 0  
 CALCOMP PAPER (n) height of paper 1  
 CALCOMP PLOT (n) curves (E) options columns (C),(C),..., (C) versus (C) V  
     CALCOMP PLOT (n),(E) vertical scales (K) to (K), (C),(C),..., (C) vs (C) V  
     CALCOMP PLOT (n),(E),(C),..., (C) vs (C) horizontal scales from (K) to (K) V  
     CALCOMP PLOT (n),(E), ordinate (K),(K),(C) vs (C) abscissa (K),(K) V  
     CALCOMP PLOT (n) curves, (E) options (C) vs (C), (C) vs (C),..., (C) vs (C) V  
     CALCOMP PLOT (n),(E), vertical scale (K),(K), (C) vs (C),..., (C) vs (C) V  
     CALCOMP PLOT (n),(E), (C) vs (C),..., (C) vs (C) horizontal scale (K),(K) V  
     CALCOMP PLOT (n),(E) vert. (K),(K), (C) vs (C),..., (C) vs (C) horiz. (K),(K) V  
 CALCOMP SIZE (n) height of paper 1  
     CALCOMP SIZE (K) height of vertical or y-axis 1  
     CALCOMP SIZE (n) height of paper (K) height of vertical axis 2  
     CALCOMP SIZE (K) height of vertical axis, (K) length of horizontal axis 2  
     CALCOMP SIZE (n) height of paper, (K) vertical height, (K) horizontal width 3  
 CALCOMP SLOW mode 0  
 CALCOMP SPEED (n) \$ n=0 for slow mode, n=1 for fast mode 1  
 CALCOMP TAPE "L" unit 0  
 CAUCHY PLOT of column (C) 1,P  
     SENSOR EQ col (C) for values equal to (E), replace by (E), put in col (C) 4

(C)=COLUMN number; (E)=EITHER col number or constant; (K)=CONSTANT; (N)=instr. number  
 N=affect NRMAX; P=print; S=synonym; V=args variable; W=work below NRMAX; X=no storage  
 A=abbreviation; B=below NRMAX; C=cannot store; D=args odd; E=args even; M=must store;  
 (R)=ROW number; (small letter) =always integer; qualifier "L"= LETTER A,B,C,D,E or F.

CENSOR GE col (C) nos greater than or eq. to (E), replace by (E), put in (C)	4
CENSOR GT col (C) for values greater than (E), replace by (E), put in col (C)	4
CENSOR LE col (C) nos less than or equal to (E), replace by (E), put in (C)	4
CENSOR LT col (C) for values less than (E), replace by (E), put in col (C)	4
CENSOR NE col (C) for values not equal to (E), replace by (E), put in col (C)	4
CHOOSE rows with (K) in column (C) and put in column (C)	3,N
CHOOSE rows with numbers between (K) and (K) in column (C) and put in col (C)	4,N
CHOOSE rows with (K) in (C), corresp. rows of (C),..., (C) put in (C),..., (C)	E,N,V
CHOOSE from (K) to (K) in (C), corr. rows of (C),..., (C) put in (C),..., (C)	D,V,N
CODE column (C) using length (K) put in column (C)	3
CODE column (C) starting at (K) using length (K) put in column (C)	4
COSINTEGRAL of (E) put in column (C)	2
COSINTEGRAL of (E) put real values in col (C) and imaginary values in col (C)	3
CPLOT col (C), symbol (E) col (C),... symbol (E) versus column (C)	D,P
CPLOT col (C) sym (E),(C),...(E) vertical scale from (K) to (K) vs col (C)	D,P
CPLOT col (C) sym (E),(C),...(E) vs col (C) horiz. scale from (K) to (K)	D,P
CPLOT (C) (E),(C),...(E) vertical (K) to (K) vs (C) horizontal (K) to (K)	D,P
CREAD UNIT "L", using (r) logical records, into columns (C), (C), ... (C)	V,S,N
CREAD UNIT "L" "L" format, using (n) blocks, into cols (C) ... (C)	V,S,N
CSET UNIT "L", using (r) logical records, into column (C)	2,S,N
CSET UNIT "L", using (r) logical records, into row (R) of column (C)	3,S,N

\*\*\*\*\* D \*\*\*\*\*

DANSK	0
DAYS for month (E) day (E) and year (E) put in column (C)	4
DELETE rows having (K) in rows of any cols (C),..., (C) put in cols (C),..., (C)	V,D,N
DELETE rows with values between (K) and (K) in (C),..., (C) put in (C),..., (C)	E,V,N
DEUTSCH	0
DEXPONENTIAL PLOT of column (C)	1,P
DIFFERENCES of y in column (C)	1,P,X
DIFFERENCES of y in column (C) put in columns (C), (C), ... , (C)	V,P
DIVDIFFERENCES for x in col (C) and y in col (C)	2,P,X
DIVDIFFERENCES for x in col (C) and y in col (C) put in (C), (C), ... , (C)	V,P

\*\*\*\*\* E \*\*\*\*\*

EEXPINTEGRAL for (n) of (E) put in column (C)	3
EINTEGRAL of (E) put in column (C)	2
ENDFILE UNIT "L"	0,S
ENGLISH	0
ESPAÑOL	0
EXPINTEGRAL for (n) of (E) put in column (C)	3
EXPONENTIAL PLOT of column (C)	1,P
EXTREME PLOT of column (C)	1,P
EXTREME PLOT with parameter (K) of column (C)	2,P

(C)=COLUMN number; (E)=EITHER col number or constant; (K)=CONSTANT; (N)=instr. number  
N=affect NRMAX; P=print; S=synonym; V=args variable; W=work below NRMAX; X=no storage  
A=abbreviation; B=below NRMAX; C=cannot store; D=args odd; E=args even; M=must store;  
(R)=ROW number; (small letter) =always integer; qualifier "L"= LETTER A,B,C,D,E or F.

\*\*\*\*\* F \*\*\*\*\*

FOURPLOTS of (C) vs (C) and (C) vs (C) and (C) vs (C) and (C) vs (C) ... 8,P  
FOURPLOTS (C), (K), (K), (C), (K), (K), (C), (K), (K), (C), (K), (K), (C), (K), (K) E,P  
FRANCAIS 0  
FULL \$ no arguments 0

\*\*\*\*\* G \*\*\*\*\*

GAMMA PLOT with parameter (K) of column (C) 2,P

\*\*\*\*\* H \*\*\*\*\*

HALFNORMAL PLOT of column (C) 1,P  
HCOSINTEGRAL of (E) put in column (C) 2  
HCOSINTEGRAL of (E) put real part in col (C) and imaginary part in col (C) 3  
HSININTEGRAL of (E) put in column (C) 2

\*\*\*\*\* I \*\*\*\*\*

INTERACTIVE use from a terminal 0,C  
INTERACTIVE use with (c) characters per line 1,C  
ITALIANO 0

\*\*\*\*\* J \*\*\*\*\*

JAPANESE 0

\*\*\*\*\* L \*\*\*\*\*

LABEL (label,) (label,) (label,) ... V,C  
LABEL (label,) (C), (label,) (C), (label,) (C,) ... V,C  
LAMBDA PLOT with parameter (K) of column (C) 2,P  
LENGTH equal to (n) lines printed per page 1  
LOCAL printing at terminal 0  
LOGISTIC PLOT of column (C) 1,P  
LOGNORMAL PLOT of column (C) 1,P

\*\*\*\*\* M \*\*\*\*\*

MEDIAN of numbers in column (C) put in column (C) 2

(C)=COLUMN number; (E)=EITHER col number or constant; (K)=CONSTANT; (N)=instr. number  
N=affect NRMAX; P=print; S=synonym; V=args variable; W=work below NRMAX; X=no storage  
A=abbreviation; B=below NRMAX; C=cannot store; D=args odd; E=args even; M=must store;  
(R)=ROW number; (small letter) =always integer; qualifier "L"= LETTER A,B,C,D,E or F.

MEDIAN of numbers in cols (C), (C), ..., (C) put in cols (C), (C), ..., (C) E  
 MLABEL (label,) (R,) (C,) (r,) (c,), (label,) (R,) (C,) (r,) (c,), ... V,C

\*\*\*\*\* N \*\*\*\*\*

NC PLOT (C), (E), ..., (C), (E) versus column (C) D,P  
 NC PLOT (C), (C), ..., (C), (E) vertical scale (K) to (K) vs col (C) D,P  
 NC PLOT (C), (E), ..., (C), (E) vs col (C) horizontal scale (K) to (K) D,P  
 NC PLOT (C), (C), ..., (C), (E) vertical (K) to (K) vs (C) horizontal (K) to (K) D,P  
 NEDERLANDS 0  
 NEGEINTEGRAL of (E) put in column (C) 2  
 NORMAL PLOT of column (C) 1,P  
 NORSE 0  
 NPLOT columns (C), (C), ..., (C) versus column (C) V,P  
 NPLOT cols (C), (C), ..., (C) vertical scale from (K) to (K) versus col (C) V,P  
 NPLOT cols (C) ... (C) versus column (C) with horiz. scale from (K) to (K) V,P  
 NTABLE AVERAGE (n) way with levels in cols (C), (C), ..., (C) for data in (C) V,P,X  
 NTABLE AVERAGE (n) way, levels (C) ... (C), data in (C) start storing in (C) V,P  
 NTABLE FREQUENCY (n) way with levels in cols (C), (C), ..., (C) V,P,X  
 NTABLE FREQUENCY (n) way, levels in (C), (C), ..., (C) start storing in (C) V,P  
 NTABLE MAXIMUM (n) way, levels in cols (C), (C), ..., (C) for data in col (C) V,P,X  
 NTABLE MAXIMUM (n) way, levels (C) ... (C) data in (C) start storing in (C) V,P  
 NTABLE MEDIAN (n) way with levels in cols (C), (C), ..., (C) for data in col (C) V,X,P  
 NTABLE MEDIAN (n) way, levels (C), ..., (C) data in (C) start storing in (C) V,P  
 NTABLE MINIMUM (n) way with levels in cols (C), (C) ... (C) for data in col (C) V,P,X  
 NTABLE MINIMUM (n) way, levels (C) ... (C) data in (C) start storing in (C) V,P  
 NTABLE PERCENTAGE (n) way with levels in cols (C), (C), ..., (C) V,P,X  
 NTABLE PERCENTAGE (n) way, levels in (C), (C), ..., (C) start storing in (C) V,P  
 NTABLE PROPORTION (n) way with levels in cols (C), (C), ..., (C) V,P,X  
 NTABLE PROPORTION (n) way, levels in (C), (C), ..., (C) start storing in (C) V,P  
 NTABLE RANGE (n) way with levels in cols (C), (C), ..., (C) for data in col (C) V,P,X  
 NTABLE RANGE (n) way, levels (C), ..., (C) data in (C) start storing in (C) V,P  
 NTABLE STDEV (n) way with levels in cols (C), (C), ..., (C) for data in col (C) V,P,X  
 NTABLE STDDEV (n) way, levels (C) ... (C) data in (C) start storing in (C) V,P  
 NTABLE SUM (n) way with levels in cols (C), (C), ..., (C) for data in col (C) V,P,X  
 NTABLE SUM (n) way, levels (C), ..., (C) data in (C) start storing in (C) V,P

\*\*\*\*\* O \*\*\*\*\*

OMIT rows with (K) in column (C) and put in column (C) 3,N  
 OMIT rows with numbers between (K) and (K) in column (C) and put in col (C) 4,N  
 OMIT rows with (K) in col (C), corr. rows of (C), ..., (C) put in (C), ..., (C) E,V,N  
 OMIT from (K) to (K) in (C), corresp. rows of (C), ..., (C) put in (C), ..., (C) D,V,N

\*\*\*\*\* P \*\*\*\*\*

PARETO PLOT with parameter (K) of column (C) 2,P  
 PERCENTAGES of numbers on column (C) put in column (C) 2

(C)=COLUMN number; (E)=EITHER col number or constant; (K)=CONSTANT; (N)=instr. number  
 N=affect NRM<sub>AX</sub>; P=print; S=synonym; V=args variable; W=work below NRM<sub>AX</sub>; X=no storage  
 A=abbreviation; B=below NRM<sub>AX</sub>; C=cannot store; D=args odd; E=args even; M=must store;  
 (R)=ROW number; (small letter) =always integer; qualifier 'L'= LETTER A,B,C,D,E or F.

PERCENTAGES of numbers in cols (C) ... (C) put in cols (C) ... (C)	E
POISSON PLOT with parameter (K) of column (C)	2,P
PORTUGESE	0
PROPORTIONS of numbers in column (C) put in column (C)	2
PROPORTION of numbers in cols (C) ... (C) put in cols (C) ... (C)	E

\*\*\*\*\* R \*\*\*\*\*

RANGE of numbers in column (C) put in column (C)	2
RANGE of numbers in cols (C) ... (C) put in cols (C) ... (C)	E
READ UNIT "L" into columns (C), (C) ... (C)	V,N,S
READ UNIT "L" "L" format into columns (C), (C), ... (C)	V,N,S
RECODE column (C) into column (C)	2
REMOTE printing on high-speed printer	0
REPLACE the number (K) in column (C) by the number (K) and put in column (C)	4
REPLACE the numbers from (K) to (K) in col (C) by (K) and put in column (C)	5
RETAIN numbers between (K) and (K) in cols (C), ..., (C) put in cols (C), ..., (C)	E,V,N
REWIND UNIT "L"	0,S

\*\*\*\*\* S \*\*\*\*\*

SDIFFERENCES of y in column (C) put in columns (C), (C), ... , (C)	V
SDIVDIFFERENCES for x in (C) and y in (C) put in cols (C), (C), ... , (C)	V
SET UNIT "L" into column (C)	1,N,S
SET UNIT "L" starting with row (R) of column (C)	2,N,S
SKIP UNIT "L", forward (r) logical records	1,S
SININTEGRAL of (E) put in column (C)	2
SLOVENE	0
SSTEM LEAF of column (C) put scrawl in column (C)	2
SSTEM LEAF of col (C) put scrawl in col (C) and depth in col (C)	3,B
SSTEM LEAF col (C) for (i), (j), (k) and (l), put scrawl in col (C)	6
SSTEM LEAF of (C) for (i), (j), (k) and (l), put scrawl in (C) depth in (C)	7,B
STATPLOTS of column (C)	1,P
STDDEV of numbers in column (C) put in column (C)	2
STDEV of numbers in cols (C) ... (C) put in cols (C) ... (C)	E
STEM LEAF of column (C)	1,P,X
STEM LEAF of column (C) and put scrawl in column (C)	2,P
STEM LEAF of col (C), scrawl in (C) and put depth in column (C)	3,P,B
STEM LEAF of column (C) for parameter values (k), (j), (k) and (l)	5,P,X
STEM LEAF of col (C) for (k), (j), (k) and (l) put scrawl in col (C)	6,P
STEM LEAF of col (C) for (i), (j), (k) (l) put scrawl in (C) and depth in (C)	7,P,B
SVENSKA	0

\*\*\*\*\* T \*\*\*\*\*

TABLE AVERAGE (n) way with levels in cols (C), (C), ... , (C) for data in col (C)	V,P,X
TABLE AVERAGE (n) way, levels (C), ... , (C) data in (C) start storing in (C)	V,P
TABLE FREQUENCY (n) way with levels in cols (C), (C), ... , (C)	V,P,X

(C)=COLUMN number; (E)=EITHER col number or constant; (K)=CONSTANT; (N)=instr. number  
 N=affect NRMAX; P=print; S=synonym; V=args variable; W=work below NRMAX; X=no storage  
 A=abbreviation; B=below NRMAX; C=cannot store; D=args odd; E=args even; M=must store;  
 (R)=ROW number; (small letter) =always integer; qualifier "L"= LETTER A,B,C,D,E or F.

TABLE FREQUENCY (n) way, levels in (C), (C), ..., (C) start storing in (C)	V,P
TABLE MAXIMUM (n) way with levels in cols (C), (C), ..., (C) for data in col (C)	V,P,X
TABLE MAXIMUM (n) way, levels (C), ..., (C) data in (C) start storing in (C)	V,P
TABLE MEDIAN (n) way with levels in cols (C), (C), ..., (C) for data in col (C)	V,X,P
TABLE MEDIAN (n) way, levels (C), ..., (C) data in (C) start storing in (C)	V,P
TABLE MINIMUM (n) way with levels in cols (C), (C), ..., (C) for data in col (C)	V,P,X
TABLE MINIMUM (n) way, levels (C), ..., (C) data in (C) start storing in (C)	V,P
TABLE PERCENTAGE (n) way with levels in cols (C), (C), ..., (C)	V,P,X
TABLE PERCENTAGE (n) way, levels in (C), (C), ..., (C) start storing in (C)	V,P
TABLE PROPORTION (n) way with levels in cols (C), (C), ..., (C)	V,P,X
TABLE PROPORTION (n) way, levels in (C), (C), ..., (C) start storing in (C)	V,P
TABLE RANGE (n) way with levels in cols (C), (C), ..., (C) for data in col (C)	V,P,X
TABLE RANGE (n) way, levels (C), ..., (C) data in (C) start storing in (C)	V,P
TABLE STDEV (n) way with levels in cols (C), (C), ..., (C) for data in col (C)	V,P,X
TABLE STDDDEV (n) way, levels (C), ..., (C) data in (C) start storing in (C)	V,P
TABLE SUM (n) way with levels in cols (C), (C), ..., (C) for data in col (C)	V,P,X
TABLE SUM (n) way, levels (C), ..., (C) data in (C) start storing in (C)	V,P
TAPE with (c) characters per logical record	1,C
TAPE with (c) characters/logical record and (r) logical records/block	2,C
TERMINAL \$ no arguments	0
TWOLOTS of column (C) versus column (C) and column (C) versus column (C)	4,P
TWOLOTS col (C) (K) to (K) vs (C) (K) to (K) and (C) (K) (K) vs (C) (K) (K)	E,V,P

\*\*\*\*\* U \*\*\*\*\*

UNIFORM PLOT of column (C)	1,P
UNIT with (c) characters per logical record	1,C
UNIT with (c) characters/logical record and (r) logical records/block	2,C

\*\*\*\*\* V \*\*\*\*\*

VOCABULARY	0,P
------------	-----

\*\*\*\*\* W \*\*\*\*\*

WEIBULL PLOT with parameter (K) of column (C)	2,P
WIDTH set to a maximum of (c) characters per line	1
WRITE UNIT "L" from columns (C), (C) ... (C)	V,S
WRITE UNIT "L" "L" format, from columns (C), (C) ... (C)	V,S

\*\*\*\*\* Y \*\*\*\*\*

YUGOSLAV	0
----------	---

(C)=COLUMN number; (E)=EITHER col number or constant; (K)=CONSTANT; (N)=instr. number  
N=̄affect NRMAX; P=̄print; S=̄synonym; V=̄args variable; W=̄work below NRMAX; X=no storage  
A=̄abbreviation; B=̄below NRMAX; C=̄cannot store; D=̄args odd; E=̄args even; M=̄must store;  
(R)=̄ROW number; (small letter)=̄always integer; qualifier "L"= LETTER A,B,C,D,E or F.

U.S. DEPT. OF COMM. BIBLIOGRAPHIC DATA SHEET	1. PUBLICATION OR REPORT NO. NBSIR 77-1276	2. Gov't Accession No.	3. Recipient's Accession No.
4. TITLE AND SUBTITLE  OMNITAB II User's Reference Manual 1977 Supplement		5. Publication Date July 1977	
		6. Performing Organization Code NBSIR 77-1276	
7. AUTHOR(S) David Hogben and Sally T. Peavy		8. Performing Organ. Report No.	
9. PERFORMING ORGANIZATION NAME AND ADDRESS  NATIONAL BUREAU OF STANDARDS DEPARTMENT OF COMMERCE WASHINGTON, D.C. 20234		10. Project/Task/Work Unit No.	
		11. Contract/Grant No.	
12. Sponsoring Organization Name and Complete Address (Street, City, State, ZIP)  National Highway Traffic Safety Administration Department of Transportation Washington, D. C. 20590		13. Type of Report & Period Covered Final	
		14. Sponsoring Agency Code	
15. SUPPLEMENTARY NOTES			
<p>16. ABSTRACT (A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here.)</p> <p>The Supplement describes all the additions and improvements made to the OMNITAB II computing system at NBS since 1970. OMNITAB 1977, as OMNITAB II is now known, is fully interactive. Major new capabilities now exist in the following areas: Use of labels, table making, plotting, numerical analysis, editing of data, stem-and-leaf displays, selection of variables in linear regression and probability plotting. The Supplement is an interim document for use until OMNITAB 1977 is released and new documentation prepared.</p>			
17. KEY WORDS (six to twelve entries; alphabetical order; capitalize only the first letter of the first key word unless a proper name; separated by semicolons) Automatic printing; Calcomp plotting; data editing; interactive; labels; numerical analysis; OMNITAB II computing system; probability plotting; selection of variables in linear regression; statistical plotting; stem-and-leaf displays; table making.			
18. AVAILABILITY <input checked="" type="checkbox"/> Unlimited  <input type="checkbox"/> For Official Distribution. Do Not Release to NTIS  <input type="checkbox"/> Order From Sup. of Doc., U.S. Government Printing Office Washington, D.C. 20402, SD Cat. No. C13  <input checked="" type="checkbox"/> Order From National Technical Information Service (NTIS) Springfield, Virginia 22151		19. SECURITY CLASS (THIS REPORT)  UNCL ASSIFIED	21. NO. OF PAGES  175
		20. SECURITY CLASS (THIS PAGE)  UNCLASSIFIED	22. Price  \$6.75